جامعہ ملیہ اسلامیہ

ڈاکٹر ذاکر حسین لائبریری

DR. ZAKIR HUSAIN LIBRARY

JAMIA MILLIA ISLAMIA
JAMIA NAGAR

NEW DELHI

CALL NO.

Accession No.

Call No.............. *Acc. No*....................

Founded in 1887 by G. Stanley Hall

THE American Journal OF Psychology

Edited by
Lloyd G. Humphreys   Leo J. Postman
William E. Kappauf   Norman E. Spear

March 1977 · Vol. 90, No. 1

The American Journal of Psychology

Editors
Lloyd G. Humphreys   Leo J. Postman
William E. Kappauf   Norman E. Spear
Carl P. Duncan, Book reviews   E. R. Hilgard, Necrologies

The *American Journal of Psychology* was founded in the interest of general experimental psychology and publishes reports of original research, articles focusing on important experimental problems, shorter notes and discussions, and pertinent book reviews. Manuscripts of articles and notes may be submitted to Drs. Humphreys or Kappauf (University of Illinois at Urbana-Champaign, Champaign, Illinois 61820), Spear (State University of New York, Binghamton, New York 13901), or Postman (Institute for Human Learning, University of California, Berkeley, California 94720). Books intended for review should be sent to Dr. Duncan (Northwestern University, Evanston, Illinois 60201).

The subscription price is $15.00 a year (foreign, $15.50); single copies cost $5.00. Back numbers and volumes may be obtained from the University of Illinois Press, Urbana, Illinois 61801. All business inquiries (subscriptions, advertisements, changes of address) should also be addressed to the Press.

Published quarterly, in March, June, September, and December
Office of publication, University of Illinois Press, Urbana, Illinois 61801
Second-class postage paid at Urbana, Illinois

© 1977 by the Board of Trustees of the University of Illinois
Manufactured in the United States of America

Contents

David E. Hogan and Thomas R. Zentall. Backward associations in the pigeon 3

Donald H. Mershon and Vincent L. Lembo. Scalar perceptions of distance in simple binocular configurations 17

Jane Kestner and Donald A. Walter. Awareness and the effect of rote rehearsal on free recall 29

John C. Schmitt and C. James Scheirer. Segregation strategies in memory scanning 39

Douglas K. Detterman. A comparison of item and position probes in short-term memory 45

Dennis Runcie and R. Michael O'Bannon. An independence of induced amnesia and emotional response 55

Larry G. Richards. Form class as a perceptual set 63

Frank J. Tolkmitt and Robin Brindley. Auditory perception of spatiotemporal patterns 73

H. R. Schiffman and Douglas J. Bobko. The role of number and familiarity of stimuli in the perception of brief temporal intervals 85

Robert A. Hicks, Arlene Okuda, and Dianne Thomsen. Depriving rats of REM sleep: The identification of a methodological problem 95

Jack R. Nation and John C. McCullers. An expectancy model for estimating the effects of age and incentive in probability learning 103

C. Weir. The effect of blank trials on probability learning 121

Daniel M. Platnick and Larry G. Richards. Individual differences related to performance on two word-recognition tasks 133

Edward M. Brussell, Stephen R. Stober, and Deborah M. Bodinger. Sensory information and subjective contour 145

John H. Mueller and Sam C. Brown. Output interference and intralist repetition in free recall 157

Book reviews (William F. Battig, Carl P. Duncan, Alan S. Brown, Allan Newell, Sharon D. Herzberger, Donald E. Parker 165

Consultants

Scheduled for publication

Contributors to Volume 89

AMERICAN JOURNAL OF PSYCHOLOGY
March 1977, Vol 90, No. 1, pp. 3–15

Backward associations in the pigeon

David E. Hogan and Thomas R. Zentall
University of Kentucky

The learning of backward associations by pigeons during training of forward associations was studied in three experiments using a symbolic matching task. When the sample stimulus remained present with the comparison stimuli, no evidence of learning of backward associations was found with colors as comparison stimuli and either colors (Experiment I) or shapes (Experiment II) as sample stimuli. When the sample stimulus was removed on presentation of the comparison stimuli (Experiment III), evidence of backward associations was found, but only over the first few transfer trials. The data are contrasted with the strong evidence of learning of backward associations by humans.

An association between two events may consist of two components, a forward association and a backward association. After training, evidence of the establishment of a forward association comes from an organism's ability to anticipate a second event, or its consequence, given a first event. Evidence of the establishment of a backward association comes from an organism's ability to anticipate the first event, given the second event. Asch and Ebenholtz (1962) have argued that humans learn backward associations simultaneously with forward associations and that the strength of the backward association is equal to that of the forward association. This position, called the associative-symmetry hypothesis, views an association as a bond tying the two events together.

A more traditional view argues that the backward association is established incidentally, separately, and with considerably less strength than the forward association (Cason, 1926). An even stronger statement of the unidirectionality of association comes from the position that an association is really a predictive temporal relation (Rescorla, 1967). According to this view, the first event predicts the second event, but the second event does not predict the first event; in fact, the second event predicts the absence of the first event.

Although support for the associative-symmetry hypothesis has been mixed, there is no evidence from the research with humans and verbal

materials to indicate that the forward and backward associative components are learned or forgotten independently (e.g., Battig and Koppenaal, 1965; Petrich, 1970). Humans apparently show some learning of backward associations when they learn forward associations. The establishment of backward associations has been studied in research with nonverbal organisms too but has focused on the very *existence* of backward associations after training in the forward direction.

For example, Dorcus (1932) and Bunch and Lund (1932) found that rats learned a multiple T-maze in the backward direction faster than they originally learned it in the forward direction. But in both cases, the results could be attributed to general learning to learn produced by nonspecific experience or to sequences of turns possibly common to both the forward and backward directions.

Maier (1932) used a simpler, triangular maze and also found some evidence of backward associative learning. But because his maze was elevated, the rats might have learned to run to a particular location in the experimental room as a result of prior placement in the relevant goal area.

Gray (1966) and Rodewald (1974) both studied the acquisition of backward associations in the pigeon. Gray (1966) trained three pigeons with a zero-delay symbolic matching task using colors as stimuli. A sample stimulus (one of two colors) was presented on the center of three response keys. Two colors different from that of the sample stimulus were used as comparison stimuli on the side keys. On nonreinforced test trials, the colors of the sample and comparison stimuli were interchanged. On the first test (28 trials), two of the three birds showed a preference for the color on a side key that was the backward association of the color on the center key, and all three birds showed that preference on the second test.

Rodewald (1974), using a simultaneous symbolic matching task (sample stimulus remains on with comparison stimuli), trained three pigeons with colors as sample stimuli and shapes as comparison stimuli and then reversed sample and comparison stimuli during the test, reinforcing responses to the backward associations. Although all three birds performed somewhat above chance during the extended test session, such performance might have been the result of learning during the test session rather than of transfer from the training sessions. Rodewald argued that the significant drop in performance from the last training session to the test session indicated that pigeons did not learn stimulus equivalence (e.g., the pigeons did not learn that the red key was interchangeable with the key that had a vertical line). The notion of stimulus equivalence is similar to that of associative symmetry in that both involve the interchangeability of stimuli.

EXPERIMENT I

The purpose of Experiment I was to assess the establishment of backward associations in pigeons after symbolic matching training, using a measure of learning rather than of test performance. Backward associative strength can be assessed by contrasting the rate of learning a new task in which the backward association should facilitate performance with the rate of learning a new task in which the backward association should retard performance. Such a design provides an extremely sensitive, relative measure of backward associative strength. It also allows for the appearance of transfer even if there should be an initial disruption of performance due to the novelty of the new configurations, and it still permits examination of the test data, which can be derived from the learning data by analyzing performance on early trials.

METHOD

Subjects

Twelve experimentally naive, female White Carneaux pigeons approximately a year old were the subjects. They were maintained at 80% of their free-feeding weights with water and grit available in their home cages.

Apparatus

A standard three-key Lehigh Valley Electronics pigeon test chamber was used (model 1519D). Behind each of the three horizontally aligned pecking keys was an in-line readout projector that could illuminate the key with any of four colored fields produced by Kodak Wratten Filters: red (26), green (60), blue (38A), or yellow (9). Electromechanical control equipment was located in an adjacent room. Sounds were masked and ventilation provided by a blower fan.

Procedure

On the first day, all birds were magazine trained and shaped to respond to a singly lit side key. Shaping position (left or right) and color (blue and yellow) were counterbalanced over birds. On the second day, each bird was given 48 continuous reinforcements for responding to each side-key color presented equally often on the left and right. On the third day, all birds were introduced to the symbolic matching task involving four possible configurations: red or green on the center key, either with blue on the left and yellow on the right key or with yellow on the left and blue on the right key. Each configuration was presented once in each four-trial block. Neither color and neither side-key position was correct on more than three consecutive trials. Each session consisted of 96 trials. During the 16 days of this training (phase 1), half the birds were reinforced for responding to the blue side key when the center key was green and to the yellow side key when the center key was red. The remaining birds were reinforced for responding to the blue side key when the center key was red and to the yellow side key when the center key was green.

Trials began with the onset of the house light and the center key. Five pecks to the center key illuminated all three keys. A single response to either side key terminated the trial: a correct response was followed by 1.5-sec access to mixed grain (Purina pigeon chow) and a 5-sec intertrial interval; an incorrect response resulted in a 5-sec intertrial interval.

After phase 1, half the birds, *group P*, were assigned to the positive transfer condition; the other half, *group N*, were assigned to the negative transfer condition. The groups were matched for phase 1 performance. On the day after the last session of phase 1, all birds were transferred to a set of configurations in which blue or yellow appeared on the center key and red and green appeared on the side keys (phase 2). As during phase 1, these configurations were presented in four-trial blocks, with each configuration represented once in each block. For group P, reinforcement was delivered after a response to the side key that was the backward associate of the center key; that is, if during phase 1 responses to blue on the side key were reinforced when the center key was red, then during phase 2 responses to red on the side key were reinforced when the center key was blue. For group N, a response to the other side key was reinforced; that is, if during phase 1 responses to blue on the side key were reinforced when the center key was red, then during phase 2 responses to green on the side key were reinforced when the center key was blue. Birds were given 13 daily sessions of training with the phase 2 configurations. Again, each session consisted of 96 trials.

RESULTS AND DISCUSSION

Acquisition data for the forward associations are presented in Figure 1. Transfer data are presented in Figure 2. First-session transfer results indicated no difference between performance of group P and group N [$F < 1$], with both groups performing at chance. Gray (1966) found a preference for the backward association over 28 test trials, but in the present experiment, little difference in performance between group P and group N was found over the first 28 transfer trials [$F < 1$]. In addition, there was no evidence that group P birds acquired the transfer task faster than group N birds by either of two criteria: sessions to 60% correct [$F(1, 10) = 3.30, p > .05$] or sessions to 80% correct [$F < 1$]. For all subsequent analyses, the .05 level of significance was adopted.

It is possible that differences in transfer were not observed because eight of the twelve birds demonstrated strong color preferences when first exposed to the new configurations. On the first transfer session, two group P birds and four group N birds showed a consistent tendency to respond to the red side key regardless of the color of the center key, and two group P birds consistently selected the green side key. The four other animals showed complete position responding during the first transfer session. Because perseverative stimulus responding might have masked differential transfer effects, at least during early transfer trials, a second experiment was carried out.

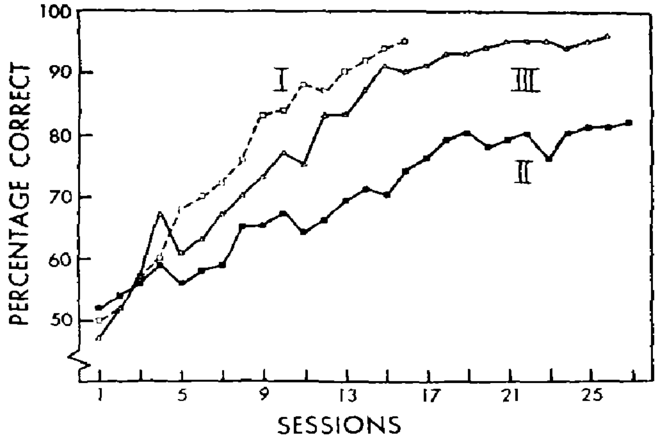


Figure 1. Acquisition of symbolic matching in Experiments I, II, and III. Experiment I: simultaneous task; sample and comparison stimuli were colors; five required responses to sample. Experiment II: simultaneous task; sample and comparison stimuli were shapes and colors; five required responses to sample. Experiment III: zero-delay task; sample and comparison stimuli were colors; ten required responses to sample

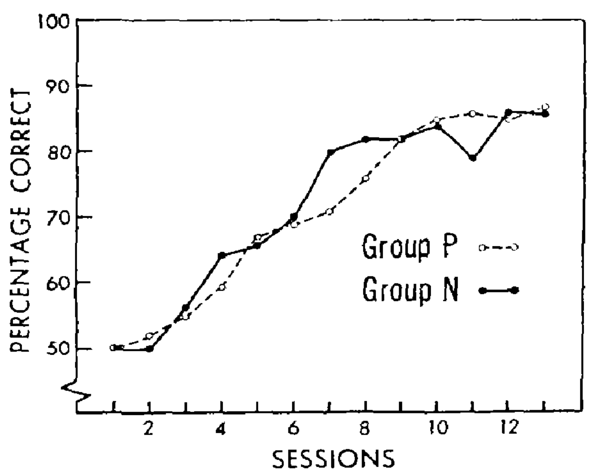


Figure 2. Acquisition of transfer task for positive and negative transfer groups in Experiment I

EXPERIMENT II

The second experiment was an attempt to reduce the stimulus preferences that might have obscured transfer differences in Experiment I. Two shapes were substituted for two of the colors used as stimuli in

Experiment I. In unpublished research we have found that preferential responding occurs much less in similar tasks (matching and oddity) with shapes than with colors as stimuli.

In all other respects, the design of Experiment II was the same as that of Experiment I.

METHOD

Subjects

Twelve experimentally naive, female White Carneaux pigeons were the subjects. They were maintained at 80% of their free-feeding weights throughout the experiment.

Apparatus and procedure

The apparatus was the same as that used in Experiment I. A white annulus (18 mm outside diameter, 11 mm inside diameter) on a black background and a white cross (consisting of a vertical and a horizontal line each 18 mm long and 3.5 mm wide) on a black background were added to the stimulus projectors. Phase 1 configurations involved shapes on the center key and colors (red and green) on the side keys. Birds were trained for 27 daily sessions (phase 1), after which the phase 2 transfer sessions began. As in Experiment I, the sample and comparison stimuli were interchanged during phase 2. Phase 2 consisted of 12 daily sessions. In both phases again, each session consisted of 96 trials.

RESULTS AND DISCUSSION

Results of the forward associative training of Experiment II are presented in Figure 1. Two birds were dropped at the end of phase 1 for failure to acquire the forward associations. The shape/color problem was learned significantly slower than the color/color problem in Experiment I [$F(1, 22) = 19.50$]. This result is consistent with Carter and Eckerman's (1975) finding that a symbolic matching task involving shapes and colors as sample and comparison stimuli is learned slower than a matching task with colors alone.

Phase 2 transfer data are presented in Figure 3. As in Experiment I, the phase 2 performance of group P did not differ significantly from that of group N over the first 28 trials; nor did it differ over the first 96 trials [both Fs < 1]. All birds reverted to position preferences during the first transfer session. Sessions to 60% correct indicated better performance for group P than for group N, but the difference was not significant [$F(1, 8) = 2.34$]. The difference in performance to a criterion of sessions to 80% correct was not significant either [$F(1, 8) = 4.06$]. What appears to be slower learning by group N in Figure 3 can be attributed largely to one bird that performed well below the others in that group.

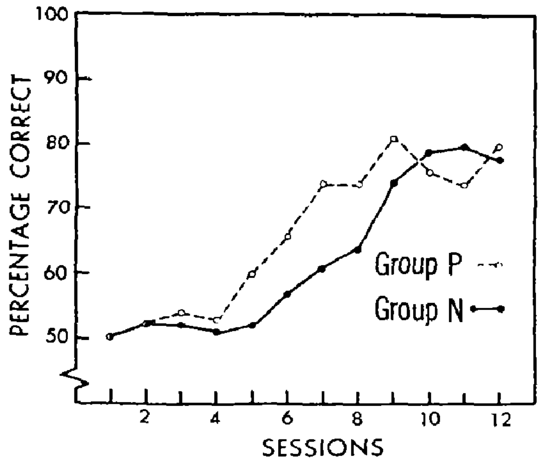


Figure 3. Acquisition of transfer task for positive and negative transfer groups in Experiment II

Thus, Experiments I and II did not support Gray's (1966) conclusion that backward associations are established during training on a symbolic matching task. When color/color associations were used in Experiment I, strong color preferences were found for most birds on the first transfer session, and later transfer sessions indicated nondifferential acquisition rates. In an attempt to reduce the likelihood that stimulus preferences would mask transfer, Experiment II employed shapes and colors. As in Experiment I, the data on the first transfer session indicated chance responding for most animals. Although these data might have been influenced by initial stimulus preferences, position preferences, or nonsystematic responding produced by the stimulus change, in neither experiment was there evidence that group P birds abandoned perseverative responding sooner than group N birds, and though in Experiment II group P appeared to learn the second task faster than group N, the difference was not significant.

It is possible that the failure to observe backward associative learning in Experiments I and II was due to the fact that a 'simultaneous' task was used. In the simultaneous task, reinforcement was provided, for example, for a response to the blue side key in the presence of a red center key. The backward-association configuration in Experiment I, phase 2, involved blue on the center key with red and green on the side keys. During the early trials of phase 2, birds were observed to peck the center key, move to the appropriate backward-association stimulus but not peck, then return and peck the center key. Such behavior would often

ɔntinue for a hundred or more pecks to the center key before the birds 'ould peck the side key. Because the birds observed the side key and ıen pecked the center key, they continued, in effect, to respond to the ›rward association they had learned. Perhaps the lack of evidence of ackward associations was due to nonreinforced responses to the center ey that tended to extinguish the forward associations (and perhaps the ackward associations as well) during the early trials of phase 2. Fortui›us extinction of the backward association may be made less probable y using a procedure in which the stimulus on the center key is removed 'hen those on the side keys are presented (zero delay).

XPERIMENT III

In Experiment III a zero-delay symbolic matching task was used. Two ıeasures of backward associative strength were examined: (a) transfer ›, and acquisition of, backward-association configurations compatible r incompatible with prior forward-association training, as in Experiments and II; and (b) loss of forward associative strength after acquisition of ɔmpatible or incompatible backward associations. If backward and forard associations depend on one another for their strength, one would ‹pect some loss of the forward association for a negative-transfer group ıat learned an incompatible backward association and no loss of the ›rward association for a positive-transfer group that learned a compatible ackward association. If the stimulus change introduced at the start of hase 2 should disrupt performance and mask transfer effects, then loss f the forward association produced by acquisition of an incompatible ackward association may be a more sensitive measure of transfer.

IETHOD

ubjects

Twelve experimentally naive, female White Carneaux pigeons were the subcts. They were maintained at 80% of their free-feeding weights throughout e experiment.

pparatus and procedure

The apparatus, design, and procedure were the same as in Experiment I cept for the following changes. A zero-delay symbolic matching task was emoyed, in which ten responses to the center key terminated that stimulus and nultaneously produced the stimuli on the side keys. The number of pecks to e center key required to produce the stimuli on the side keys was increased ɔm five to ten in order to facilitate acquisition of the task. Sacks, Kamil, and ack (1972) have found that the rate of learning a zero-delay matching-to-

sample task increased as the number of required responses to the center key increased. The present experiment was conducted in four phases. Phase 1 involved acquisition of the forward associations (26 daily sessions). In phase 2, the birds were transferred to new configurations for which a response to the backward association was reinforced for group P and nonreinforced for group N (14 daily sessions). Phase 3, a test for loss of forward associative strength, involved a return to phase 1 forward-association configurations for one session. Phase 4 involved three sessions of phase 1 and 2 configurations randomly intermixed, the assumption being that sessions with mixed forward- and backward-association trials should be maximally disruptive for group N. Again, each session consisted of 96 trials.

RESULTS AND DISCUSSION

Acquisition data for forward associations trained with the zero-delay procedure of Experiment III are presented in Figure 1. The zero-delay task using colors as stimuli was learned significantly faster than the simultaneous task using a combination of colors and shapes as stimuli in Experiment II [$F(1, 22) = 5.99$]. One might have expected the zero-delay task to have been learned slower than the simultaneous task, but the zero-delay task involved the easier color/color associations (Carter and Eckerman, 1975) and was probably further facilitated by the larger number of required responses to the center key.

The difference between phase 1 performance in Experiment I (simultaneous symbolic color matching) and Experiment III (zero-delay symbolic color matching) was not significant [$F(1, 22) = 2.89$], but the difference in tasks might have been partially obscured by the difference in required responses to the center key.

Phase 2 (transfer) data for groups P and N are presented in Figure 4. No apparent differences in transfer were observed over the 13 transfer sessions. An analysis of the data on the early trials of the first transfer session indicated a significant difference over the first 28 trials [$F(1, 10) = 7.73$], but not over 48 trials [$F(1, 10) = 4.24$] or 96 trials [$F(1, 10) = 2.33$]. Thus, initial transfer differences dissipated rapidly over reinforced practice trials. The data on the first transfer session are plotted in blocks of 16 trials in Figure 5.

Performance on forward-association configurations during phase 3 for groups P and N is presented in Figure 4. No performance decrement was observed for either group over 96 practice trials [both Fs < 1]. Early trials were also analyzed to assess temporary differential disruption, but differences were not significant over the first 4 trials or over the first 28 trials [all Fs < 1]. Apparently, learning an incompatible backward association did not retroactively interfere with the forward association.

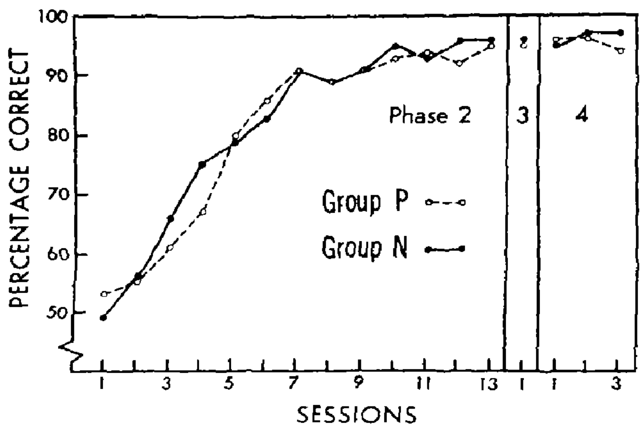


Figure 4. Acquisition of transfer task for positive and negative transfer groups in Experiment III

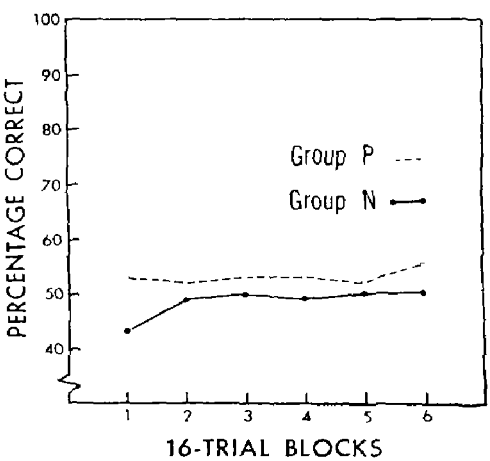


Figure 5. First session of transfer performance for positive and negative transfer groups in Experiment III

Phase 4 data also appear in Figure 4 and reflect performance on the forward-association configurations of phase 1 intermixed with the backward-association configurations of phase 2. Intermixing incompatible associations for group N did not produce a significant decline in overall performance compared with a group experiencing compatible forward and backward associations within the same practice session [$F < 1$].

Thus, the results of Experiment III, phase 2, confirm Gray's findings that pigeons trained on a zero-delay symbolic matching task show some tendency to learn backward associations. The strength of the backward

association is apparently minimal, however. Although Gray reported performance significantly better than chance over 28 test trials, the results of the present experiment suggest that the effect may dissipate within 48 trials. Moreover, the results of the tests of interference in Experiment III, phases 3 and 4, show that learning an incompatible backward association does not significantly disrupt performance of a well-learned forward association. All birds maintained high accuracy on the forward-association trials, even when incompatible backward-association trials occurred within the same practice session.

It may be that significant disruption of the forward associations could have been found for group N (relative to group P) after phase 2 of Experiment III if original training had been carried out to a less stringent criterion (e.g., 75% correct). Had a less stringent criterion been used, not only might performance have been disrupted for group N as a result of learning an incompatible backward association, but there would have been room for facilitated performance by group P (i.e., performance during phase 3 might have been significantly better than performance at the end of phase 1) as a result of learning a compatible backward association. Further research might consider manipulation of the criterion for original learning.

One might also argue that pigeons can learn backward associations while learning forward associations but that the present design was not sufficiently sensitive to assess it. As already stated, however, the comparison of a positive-transfer group with a negative-transfer group should maximize the likelihood of observing effects of backward associative learning. Thus, the evidence from the present experiments suggests that backward associations in the pigeon are weak after forward associative training and may only be observed under ideal conditions (e.g., early during the first transfer session of a zero-delay symbolic matching task).

GENERAL DISCUSSION

The development of strong backward associations has been clearly demonstrated in humans. Some learning of backward associations invariably occurs when verbal items are used as stimuli. Studies in which rats learned a maze in the reverse direction from that of original learning have confounded specific associative factors with nonspecific practice effects and so do not provide strong evidence for backward associative learning. And studies with pigeons have indicated that forward associative training results in at best weak, and easily modified, backward associations.

The failure to find evidence of strong backward associations by pigeons

may be attributable to the many procedural differences between experiments with humans and pigeons, among which two stand out. First, humans must typically *recall* the stimulus that goes with each response item (the stimulus is not present), whereas the pigeons are presented both a response item and a choice of two stimuli and must merely *recognize* the appropriate stimulus (the correct stimulus is present). But studies with humans that have used a recognition task (Cheung and Goulet, 1968; Leicht and Kausler, 1965) have found evidence of the development of strong backward associations. Thus, the evidence on backward associations by pigeons is probably not related to the method of testing. Second, humans, but not pigeons, are given instructions to identify the item that was originally paired with each of the items they are about to see. But given the forced-choice nature of the task for pigeons, the strength of an association in the backward direction should determine the extent to which a response is made to the backward associate independent of instructions to do so.

From the evidence presented here one might conclude that the learning of backward associations is related to the cognitive capacities of the species, but it is also possible that the development of backward associations depends upon the species-specific functional value of such associations (i.e., humans may *need* to be able to develop backward associations whereas pigeons may not). Such a distinction between cognitive capacity and functional value requires the identification of a species that is low in intellectual capacity but for which backward associations have high utility. Suggestions are welcomed.

Notes

This research was supported by Grants MH19757 and MH24092 from the National Institutes of Mental Health to the second author, and it is based on a master's thesis submitted to the University of Pittsburgh by the first author. Offprints may be obtained from either author, Department of Psychology, University of Kentucky, Lexington, Kentucky 40506. Received for publication October 14, 1975.

References

Asch, S. E., and Ebenholtz, S. M. 1962. The principle of associative symmetry. *Proceedings of the American Philosophical Society* 106:125–163.

Battig, W. F., and Koppenaal, R. J. 1965. Associative asymmetry in S-R vs. R-S recall of double-function lists. *Psychological Reports* 16:287–293.

Bunch, M. E., and Lund, W. R. 1932. An experiment on backward associations in animal learning. *Journal of Comparative Psychology* 13:143–156.

Carter, D. E., and Eckerman, D. A. 1975. Symbolic matching by pigeons: Rate of learning complex discriminations predicted from simple discriminations. *Science* 187:662–664.

Cason, H. 1926. Specific serial learning; a study of backward associations. *Journal of Experimental Psychology* 9:195–227.

Cheung, C. C., and Goulet, L. R. 1968. Retroactive inhibition of R-S associations in the A-B, B-C, C-B paradigms. *Journal of Experimental Psychology* 76:321–322.

Dorcus, M. D. 1932. The establishing of backward associations in forward learning of the maze by albino rats. *Journal of Comparative Psychology* 8:11–17.

Gray, L. 1966. Backward association in pigeons. *Psychonomic Science* 4:333–334.

Leicht, K. L., and Kausler, D. H. 1965. Functional stimulus learning as related to degree of practice and meaningfulness. *Journal of Experimental Psychology* 69:100–101.

Maier, N. R. F. 1932. A study of orientation in the rat. *Journal of Comparative Psychology* 14:387–399.

Petrich, J. A. 1970. S-R and R-S unlearning as a function of transfer paradigm. *Journal of Experimental Psychology* 83:19–24.

Rescorla, R. A. 1967. Pavlovian conditioning and its proper control procedures. *Psychological Review* 74:71–80.

Rodewald, H. K. 1974. Symbolic matching-to-sample by pigeons. *Psychological Reports* 34:987–990.

Sacks, R. A., Kamil, A. C., and Mack, R. 1972. The effects of fixed-ratio sample requirements on matching to sample in the pigeon. *Psychonomic Science* 26:291–293.

Consultants

The editors of the *American Journal of Psychology* take this opportunity to thank those who have recently assisted in the evaluation of manuscripts.

Jack Adams
Matthew Alpern
Fred Attneave
Michael Birnbaum
Ann Brown
Donald S. Ciccone
George Collier
Clyde H. Coombs
Don E. Dulaney
Alice Hendrickson Eagly
Selby H. Evans
Daniel Fallon
Samuel Fillenbaum
Martin E. Fishbein
William Fishbein
Ronald J. Gandelman
Walter Gogel
David Gordon
William C. Gordon
Robert Henderson
James E. Hoffman
Jerry Homzie
Earl Hunt
Larry L. Jacoby
Lawrence Jones
Alan C. Kamil
David Klahr
Stephen M. Kosslyn
Richard V. Krane
Patrick Laughlin
Dale Leonard
Charles Lewis
Stephen Lisman
Robert S. Lockhart
Lawrence Marks
Jana Lucas Mason
William J. McGuire
George Miller
Murray S. Miron
Charles E. Osgood
Peter G. Polson
Michael Posner
David Riccio
Eleanor Rosch
Jerry Rudy
Stanley R. Scobie
Keith G. Scott
Roger N. Shepard
D. A. Sommers
David Sparks
Robert J. Sternberg
Constantine Trahiotis
William S. Verplanck
Wayne Wickelgren
Paul Wong

AMERICAN JOURNAL OF PSYCHOLOGY
March 1977, Vol. 90, No. 1, pp. 17–28

Scalar perceptions of distance in simple binocular configurations

Donald H. Mershon and Vincent L. Lembo
North Carolina State University

Under reduced viewing, a single visual object tends to appear about 2 m away. Gogel (1972) found that the far point in a binocular configuration of two points of light also tended to appear about 2 m away, while the near point was displaced toward the observer. Attempts to replicate this latter result with points of light, as well as with a large rectangle for the nearer object, proved generally successful. Under some conditions, however, the far point was perceived to be closer to the observer when seen with the near point than when seen alone or with the rectangle. This unexpected result suggests caution in assuming that the far object in such a configuration will always remain at a stable perceived distance.

An individual object presented in an otherwise completely dark visual surround and with minimal cues to its egocentric distance will usually appear about 2 m away, regardless of the particular shape or size of the object. This characteristic result is what Gogel (1969) has termed the *specific-distance tendency*. The tendency may be used to explain the systematic occurrence of scalar perceptions of extent under reduced viewing (see Gogel, 1973a, 1973b). Since such reduction is seldom complete, however, it is necessary to consider the effect of whatever minimal cues remain.

For example, it is probable that the oculomotor cues to distance cannot be completely removed from any visual situation. Although it is usually possible to maintain a constant value of accommodation or convergence or to make the state of accommodation irrelevant to clarity (as by the use of pinholes), these operations are not necessarily equivalent to eliminating all cues, since the visual system must still assume *some* specific state for each display. It merely becomes the case that the state it assumes need no longer be directly related to the physical characteristics of the displays. Indeed, under some conditions of reduced viewing, the eyes probably remain near an individually determined 'resting' state of intermediate value (see Leibowitz and Owens, 1975; Owens and Leibowitz, 1976).

Whether or not the oculomotor states are systematically related to the distance of a display, they likely form irreducible residual cues that can modify the perception of distance expected from the specific-distance tendency. Gogel (1972) has termed the joint function of the specific-distance tendency and the residual cues the *egocentric reference distance.* It is this reference distance, and not the specific-distance tendency alone, that provides the metric for perceived distance in reduced viewing. One remaining question, however, is what scalar perceptions will result when a configuration of two or more objects, rather than a single object, is presented in such situations.

Logically, one could expect that the two objects in a binocular depth configuration might be seen in any of several relationships to the reference distance: for example, far object at the reference distance, near object closer; or near object at the reference distance, far object farther; or the two objects each displaced from the reference distance, but in opposite directions. Gogel (1972) has found, however, that only the first result occurs. In a configuration of two points of light that were separated by a binocular depth interval but provided the observer *no direct cues to their egocentric distances,* the far point appeared at approximately the reference distance and the near point was displaced significantly closer to the observer than when it was seen singly.

The present experiments attempted to replicate Gogel's (1972) observations with points of light and examined in addition whether the same results would be obtained if the binocularly nearer object was made visually more massive than the farther object and if the residual oculomotor cues were varied to produce different values of the reference distance. In an effort to make the nearer object perceptually more stable and thus possibly more likely to form the reference for a binocular configuration, a homogeneously illuminated rectangle was used in place of the near point; the farther object remained a point. Three experiments were conducted, the principal difference among them being the distances to the visual displays. These latter variations were expected to produce systematic differences in the residual cues. The modifications were designed to determine the generality of the major findings.

METHOD

EXPERIMENT I

Observers

The observers were 144 students from an introductory course in psychology who served to fulfill course requirements. Each had a visual acuity (corrected,

if necessary) of at least 20/30 in each eye at both near and far distances, as measured with a Keystone telebinocular, and a stereoacuity of at least 145 sec, as measured by the Keystone Multi-Stereo test. All observers were naive as to the purpose of the experiment.

Apparatus

Five different displays were used for this experiment. Each display was presented as a cut-out form in front of a small, uniformly lighted surface. The five displays were a *rectangle,* 1.25 deg wide by 2.0 deg high, located 410 cm from the observer's eyes; a circular point of light 2.7 min in diameter, the *near point,* also located at 410 cm; a similar point of light, the *far point,* located at a simulated distance of 1,230 cm from the observer; both the *near and far points,* presented simultaneously; and both the *rectangle and far point,* presented simultaneously. Each display was at eye level, and all displays were viewed binocularly. The lateral separation between objects presented simultaneously was less than 2 deg measured from center to center of the objects. Luminance of all stimuli was 1.8 mlam.

In order to produce the large simulated distance of the far point, a Polaroid stereoscope was used. Thus, the far point was created by presenting two points at 410 cm from the observer, with each point covered by oppositely oriented Polaroid filters. The observer viewed these points through a set of crossed Polaroid filters, allowing each eye to see only one of these points (although both eyes could see the rectangle or the near point, if presented). The lateral offset of the two components of the far point produced a binocular disparity of 35.8 min, which simulated a depth of 820 cm behind the physical distance of 410 cm.

All presentations of the stimuli were in an otherwise completely dark visual field. Nothing was visible except the stimulus object(s). Observations were made from inside a dark booth containing a head-and-chin rest, the crossed Polaroids, and a shutter for occluding the stimuli. Throughout the experiment, a communication system permitted the experimenter to speak with and to hear the observer. A background sound of white noise was available to mask any sounds made by the equipment before the presentation.

Procedure

After the initial visual screening, each observer was escorted to the observation booth and given instructions for the experiment. The observer then remained a minute in the dark booth before being shown the appropriate display. The 144 observers were randomly assigned to one of five groups. The 24 observers in group 1 were presented the rectangle alone. The 24 observers in group 2 were presented the near point alone. The 48 observers in group 3 were presented the far point alone.[1] The 24 observers in group 4 were presented the near and far points simultaneously, and the 24 observers in group 5 were presented the rectangle and the far point simultaneously. The positions of objects to the left and right of the midline were counterbalanced in both the single and the simultaneous presentations.

For an observer shown a single object, the task was to report verbally ("in feet or inches or in some combination of feet and inches") how far the object appeared to be from his or her eyes. For an observer shown two simultaneous objects, the task was to report verbally the apparent distance of each of the

objects. The order in which the perceived distances of the two objects were reported was counterbalanced for each simultaneous presentation. For each of the presentations, the experimenter first ascertained that the proper binocular fusion had occurred (i.e., no double images were apparent) and that the far point did indeed appear farther away when it was presented simultaneously with the near point or with the rectangle. Few observers experienced any difficulty in meeting these criteria.

EXPERIMENT II

Observers

The observers were 100 students fulfilling course requirements for an introductory psychology course. The acuity and stereoacuity criteria were the same as for Experiment I. None of the observers had been in Experiment I, and all were naive as to the purpose of the experiment.

Apparatus

The apparatus used in the present study was essentially the same as that used in Experiment I. The values for the binocular disparity of the far point (relative to the near object), visual angle of the rectangle, luminance, and lateral angular separation between objects were the same as before except that the points of light in Experiment II were 5.3 min in diameter. The principal difference between the two experiments was that for Experiment II, the physical distance of all stimuli was decreased to 205 cm. The binocular disparity of 35.8 min in the Polaroid stereoscope, therefore, resulted in a simulated distance of 307.6 cm for the far point.

Procedure

The task and instructions were also the same as those used in Experiment I. Observers were randomly assigned to one of five groups of 20 observers each. The observers in group 1 were presented the rectangle alone. The observers in group 2 were presented the near point alone. The observers in group 3 were presented the far point alone. The observers in group 4 were presented the near and far points simultaneously, and those in group 5 were presented the rectangle and far point simultaneously. The left and right positions of the stimulus objects and the order of judgments in simultaneous presentations were counterbalanced.

EXPERIMENT III

Observers

The observers were 100 students fulfilling course requirements for an introductory psychology course. The acuity and stereoacuity criteria were the same as for Experiments I and II. None of the observers had been in either previous study, and all were experimentally naive.

Apparatus

The apparatus used in the present study was essentially the same as that used

for Experiment II. The values for the binocular disparity of the far point (relative to the near object), visual angle of the rectangle, luminance, and lateral angular separation between objects were the same as before except that the far point of light in Experiment III was 3.5 min in diameter. The principal difference between Experiments II and III was that for Experiment III, use of the Polaroid stereoscope was discontinued and the far point was physically (as well as stereoscopically) positioned at 307.6 cm. The near point and the rectangle continued to be located at 205 cm.

Procedure

The task and instructions were the same as in the previous experiments. Observers were randomly assigned to one of the same five groups of 20 observers each. The left and right positions of the stimulus objects and the order of judgments in simultaneous presentations were again counterbalanced.

RESULTS

The data were the verbal judgments of the perceived distance (D') of each object for each presentation in each experiment. Such verbal judgments typically involve positively skewed distributions with the sporadic occurrence of extreme values under some conditions.[2] Figure 1 shows the far-point data from two of the presentation conditions of Experiments I, II, and III, and it indicates that the data obtained followed that pattern. Because of the sort of data involved, median values were believed to best represent the results and all statistical tests of inference were nonparametric (Mann-Whitney U test). Table 1 shows the median values and semi-interquartile ranges of perceived distance (D') for each object when presented singly and when presented simultaneously with a second object.

EXPERIMENT I

The results for the individually presented objects of Experiment I are clear. The rectangle, far point, and near point were all seen at the same median distance of 304.8 cm when presented singly. The far point when presented simultaneously with the rectangle or with the near point was also seen at approximately the same median distance. That is, the far point did not significantly shift its perceived egocentric distance, regardless of the absence or presence of either nearer object.

When the near point of Experiment I was presented simultaneously with the far point, however, the near point was perceived significantly closer to the observer than when presented singly [$p < .001$, two-tailed]. When the rectangle was presented simultaneously with the far point, it

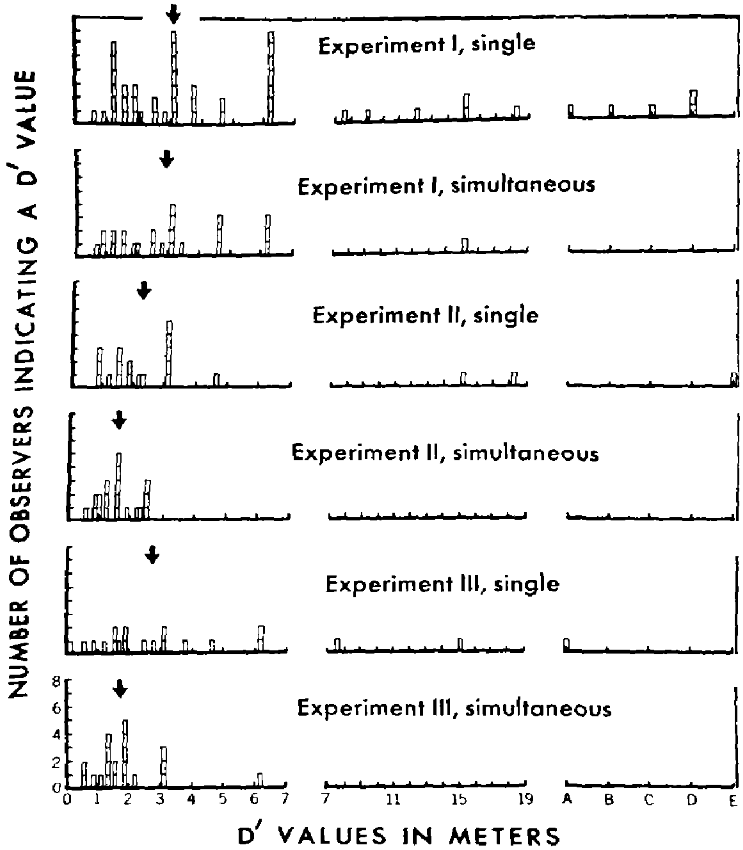


Figure 1. Distributions of the reports of perceived distance (D') of the far point when presented singly and when presented simultaneously in configuration with the near point. The horizontal axis has been segmented, in order to present the complete data in one figure without loss of detail for the majority of the values. The left and center sections are drawn to different scales, but both are in meters. The right section of the horizontal axis indicates the occurrences of extreme values without regard for scale: A, 30.5 m (100 ft); B, 45.7 m (150 ft); C, 61.0 m (200 ft); D, 1609.3 m (1 mile); and E, 'infinitely far.' The arrows indicate the median of each distribution

too was perceived significantly closer to the observer than when presented singly [$p < .01$, two-tailed]. The perceived distances of the near point and of the rectangle did not differ from each other when each was presented simultaneously with the far point. These results replicate those of Gogel (1972) and indicate in addition that the same effect occurs when a rectangle is substituted for the near point.

Table 1. Physical (P, convergence) distances, median reported values (Mdn), and semi-interquartile ranges (Q) for perceived distances of objects presented singly or simultaneously in configuration with another object; Experiments I, II, and III, all data in centimeters

	Experiment I			Experiment II			Experiment III		
	P	Mdn	Q	P	Mdn	Q	P	Mdn	Q
Single presentations									
Rectangle	410	304.8	175 3	205	274.3	240.0	205	152 4	114.3
Near point	410	304 8	95 2	205	152.4	76.2	205	175 3	217.2
Far point	1,230	304 8	213.4	308	221.0	76 2	308	259 1	190.5
Simultaneous presentations									
Near point, with far point	410	167.6	72.4	205	91 4	33.0	205	106.7	26.7
Far point, with near point	1,230	289.6	152.4	308	152 4	45 7	308	167 6	30.5
Rectangle, with far point	410	152.4	30 5	205	114 3	34 9	205	152 4	38.7
Far point, with rectangle	1,230	304 8	201.9	308	213.4	64.8	308	243.8	99.1

[a] $p > .10$ (*nonsignificant*). [b] $p < .10$, two-tailed. [c] $p < .05$, two-tailed. [d] $p < .02$, two-tailed. [e] $p < .01$, two-tailed. [f] $p < .001$, two-tailed.

EXPERIMENT II

The results of Experiment II paralleled those of Experiment I, with the following differences. First, there was more variation among the median perceived distances for single presentations. The rectangle, far point, and near point were reported at somewhat different distances, although none of the possible comparisons among these values reached significance. Second, the average perceived distances of singly presented objects were less for Experiment II than for Experiment I, although the difference was significant only for the near point [$p < .001$, two-tailed]. This outcome is in general agreement with the shorter physical distances used in Experiment II, distances which were expected to provide different values of residual cues. Third, the far point did not always appear at a constant perceived distance. Although both the rectangle and the near point were displaced toward the observer when either was presented simultaneously with the far point as compared to when either was presented singly, the far point also appeared closer to the observer when presented simultaneously with the near point than when presented singly *or* simultaneously with the rectangle [both $ps < .02$, two-tailed].

EXPERIMENT III

The reported distances to the objects presented singly in Experiment III were less than in Experiment I and not significantly different than the corresponding values in Experiment II. These results indicate that the relatively minor accommodative difference between Experiments II and III was less influential than the larger oculomotor changes between these experiments and Experiment I in determining the value of the egocentric reference distance.

The results of Experiment III for the objects presented simultaneously, while similar to those of Experiment II, are somewhat more difficult to interpret. This difficulty was due to the rather large difference in the median perceived distances of the near objects and the far point when each was presented singly. Although none of the possible comparisons among objects presented singly reached significance, the magnitude of the difference between the far point and the other objects was such that neither the rectangle nor the far point was found to shift position significantly when presented simultaneously with each other as compared to when presented singly. Thus, Experiment III contributes little further data on the effect of the rectangle on the perceived distances to the configuration.

When the near point was presented simultaneously with the far point, however, it was seen significantly closer to the observer than when presented singly [$p < .05$, two-tailed]. Also of interest is the fact that despite the large perceived distance of the far point presented singly, the far point again shifted toward the observer when presented simultaneously with the near point. This change was significant when compared to the position of the far point presented simultaneously with the rectangle [$p < .05$, two-tailed] but did not quite reach significance when compared to the far point presented singly [$p < .10$, two-tailed]. Nevertheless, this change in the perceived distance of the far point is in agreement with the findings of Experiment II and indicates that the far point may not be perceptually stable independent of the configuration in which it appears.

DISCUSSION

The experimental interests in the present research were threefold. First, we expected our results to replicate those of Gogel (1972) in showing the near point to displace toward the observer when presented simultaneously with a binocularly farther point. Second, we were interested in whether the massiveness of the rectangle would make it perceptually more stable than the point and hence more likely to be seen as the scalar reference point in a binocular configuration. Such a result — the finding that the rectangle appears at a constant distance alone or in a configuration and that the far point is displaced — was obtained in an earlier study (Mershon and Lembo, 1974) under conditions similar to those of Experiments II and III. Third, we wished to examine the generality of the results for different values of the egocentric reference distance.

MASSIVENESS

The data presented in this report are contrary to the prediction of any special 'stability' for the rectangle. The (near) rectangle was generally displaced toward the observer in the same way that the near point was. Its massiveness made no noticeable difference. Since both Experiments II and III included a replication of the earlier (positive) experiment and yet neither replicated the original finding, it seems safe to conclude that the earlier finding is not a reliable phenomenon. Thus, we have not demonstrated a manipulation of the characteristics of the nearer object that would make it more important in determining the perceived distances of the objects in a simple binocular configuration. Rather, the present results are in general agreement with those of Gogel (1972). The pos-

sibility remains, of course, that an even more massive object might have the predicted effect, despite the negative results obtained here. However, this does not appear to be a profitable line for further investigation at present.

VERBAL REPORTS

The failure to find a difference between the near point and the rectangle in the experiments just reported cannot be attributed to insufficient sensitivity of the verbal reports. For example, objects were singly presented at physical (or simulated) distances of 205, 308, 410, and 1,230 cm at different times in this research. The average median values of D' across all the singly presented objects at each of these distances show a systematic progression up to 4 m, although not beyond: $D' = 188.6$, 240.0, 304.8, and 304.8 cm respectively. This outcome lends credibility to our use of the verbal reports and is consistent with previous findings that convergence may serve as a weak cue to egocentric distance within a few meters of the observer (Gogel and Sturm, 1972).

Although verbal reports may have sufficient sensitivity for the problem at hand, their extreme variability can limit their suitability for some experiments. Asking for such verbal reports from naive observers under reduced viewing is virtually guaranteed to result in a wide variety of responses, regardless of whether or not there is extensive variation in the underlying perceptions. There are, however, few alternative measures with which to obtain information about perceived *scalar* distance. All the standard procedures of distance matching or magnitude estimation (relative to another visually defined distance) are nonscalar and therefore useless for answering the critical question of perceptual scale (see Gogel, 1968). Despite the difficulties involved in the use of verbal reports, they retain the advantage of providing the desired scalar information.

MODIFICATION OF FAR POINT D'

Perhaps the most interesting result of these experiments is the unexpected modification in the perceived distance to the far point when that point was presented simultaneously with the near point. As already noted, this result was not predicted. The a priori predictions included the possibilities that either or both of the objects might displace from their positions as presented singly, but this displacement was not expected to be in the same direction for both. Thus, the finding that the far object may appear *closer* when presented simultaneously with a yet nearer object

than when presented singly is surprising. This result suggests the possibility that even though the rectangle is as susceptible as the near point to the effects of a more distant object in a configuration, the objects in a configuration may not always follow a simple pattern with respect to the perceptual effects found for single objects. Note particularly that the far point only moved inward from its position as presented singly when it was presented simultaneously with the near point. This result did not occur when it was presented simultaneously with the rectangle. Nor did this result occur for the larger physical (and simulated) distances used in Experiment I and in the relevant study by Gogel (1972, Experiment II).

Needed now are further studies in which the number and arrangements of the binocular objects in depth are systematically varied. Needed too is more work on the interaction of the residual oculomotor cues, the specific-distance tendency, and perceived distance.

Notes

The authors thank Jeff Frederick for his help in collecting the data of Experiment II and in analyzing the data of Experiments I and II. This research was supported in part by a grant from the Research and Computer Committee of the School of Education of North Carolina State University. Requests for offprints should be sent to Donald H. Mershon, Department of Psychology, North Carolina State University, Raleigh, North Carolina 27607. Received for publication February 2, 1976; revision, May 26, 1976.

1. There were 48 observers presented the far point alone in order to give two alternative additional presentations to subgroups of 24 observers each. Since the additional presentations are of no interest to the present problem and since there were no differences between the judgments of the two subgroups to the initial presentation of the far point alone, the results of the two subgroups have been combined into a single larger group.

2. We believe that these extreme values under some conditions reflect one of two possible situations: either the observer is reporting a strictly cognitive judgment (i.e., 'it looks like a tiny point of light and must therefore be very far away'; see Carlson, 1960; Carlson and Tassone, 1962) or the observer genuinely perceives the light as being quite far away but is assigning a much larger number to describe that perception of 'far' than are most other observers. Some combination of these alternatives is, of course, also possible. We do not believe, however, that a report of 'a mile' should necessarily be treated as equivalent to a report of 'six and a half feet,' in terms of meaning or precision. Since we cannot determine which reports are cognitive and which are perceptual, we treat all responses as reports of perceived distance. In order to do statistical analyses of such data, we rely on nonparametric procedures for which the extreme responses are simply the largest values, without regard to their numerical extremeness. This approach has the advantage of retaining all the data, in-

cluding an occasional report that an object appears 'infinitely far away' (see Figure 1).

References

Carlson, V. R. 1960. Overestimation in size-constancy judgments. *American Journal of Psychology* 73:199–213.

Carlson, V. R., and Tassone, E. P. 1962. A verbal measure of the perspective attitude. *American Journal of Psychology* 75:644–647.

Gogel, W. C. 1968. The measurement of perceived size and distance. In *Contributions to sensory physiology*, vol. 3, ed. W. D. Neff. New York: Academic Press.

Gogel, W. C. 1969. The sensing of retinal size. *Vision Research* 9:1079–1094.

Gogel, W. C. 1972. Scalar perceptions with binocular cues of distance. *American Journal of Psychology* 85:477–497.

Gogel, W. C. 1973a. The organization of perceived space: I. Perceptual interactions. *Psychologische Forschung* 36:195–221.

Gogel, W. C. 1973b. The organization of perceived space: II. Consequences of perceptual interactions. *Psychologische Forschung* 36:223–247.

Gogel, W. C., and Sturm, R. D. 1972. A comparison of accommodative and fusional convergence as cues to distance. *Perception and Psychophysics* 11: 166–168.

Leibowitz, H. W., and Owens, D. A. 1975. Anomalous myopias and the intermediate dark focus of accommodation. *Science* 189:646–648.

Mershon, D. H., and Lembo, V. L. 1974. Relative visual massiveness and perceived egocentric distance. Paper presented at the meeting of the Psychonomic Society in Boston in November.

Owens, D. A., and Leibowitz, H. W. 1976. Oculomotor adjustments in darkness and the specific distance tendency. *Perception and Psychophysics* 20:2–9.

AMERICAN JOURNAL OF PSYCHOLOGY
March 1977, Vol. 90, No. 1, pp. 29–37

Awareness and the effect of rote rehearsal on free recall

Jane Kestner and Donald A. Walter
University of Notre Dame

The effect of time of awareness of a subsequent test of recall and the relationship of that awareness and rote rehearsal were studied by telling subjects which specific items to encode before the item's presentation (prior instructions) or after its rehearsal (postrehearsal instructions) and by varying rehearsal intervals for individual items. Rehearsal interval had little effect on recall, and only the prior instructions facilitated recall. This facilitation was at the same level at all rehearsal intervals, indicating that instructions affect an item's encoding only during its presentation and independent of rote rehearsal.

Several models of memory (Waugh and Norman, 1965; Sperling, 1967; Atkinson and Shiffrin, 1968; Norman and Rummelhart, 1970) attribute a dual function to rehearsal. These models contend that rehearsal not only maintains items in some type of short-term storage but is also responsible for their transfer into long-term storage. Direct attempts to show that the latter is the case have not been entirely satisfactory.

For example, Tulving (1966) manipulated rehearsal by having subjects repeat either the experimental items or unrelated items six times before learning for free recall began. He found the learning curves to be indistinguishable under the two conditions, suggesting that rote repetition of items is not an effective aid to recall. Glanzer and Meinzer (1967) had subjects either repeat each item after its presentation or remain silent. They found that the repetition resulted in poorer free recall of all items except those in the last three serial positions, indicating that while repetition may aid item storage in short-term memory, it interferes with long-term storage. Meunier, Ritz, and Meunier (1972) manipulated rehearsal in learning for free recall by either having subjects perform a distractor task after the presentation of each item or allowing them free time for rehearsal. Subjects were asked to recall each item after its presentation and, at the end of the experiment, to recall all previous items. While the initial recall of subjects given time for rehearsal was superior, their final recall did not differ from that of subjects for whom rehearsal was prevented.

These studies show that increasing time for rehearsal does not invariably improve long-term retention, a position in conflict with predictions from the models mentioned at the outset. This in turn has led to much speculation about subjects' ability to adopt different rehearsal strategies in light of different task requirements (Restle, 1970; Butterfield and Peltzman, 1971). Some authors hypothesize that a strategy may maximize either long-term or short-term retention, but not both (Jacoby and Bartz, 1972; Craik and Lockhart, 1972). They suggest that a subject has the option of using different levels of processing, only some of which lead to improved long-term retention, and that long-term retention is a function of the amount of time used for deeper processing.

In an attempt to allow subjects to employ such strategies freely, Meunier, Kestner, Meunier, and Ritz (1974) performed an experiment manipulating both the time available for rehearsal and the time at which the subjects were told they would subsequently be asked to recall the items. The latter was varied between subjects; they were informed of the test of recall either before the presentation of the first item or after all the items (consonant trigrams) had been presented. To make the rehearsal of the items observable, all subjects were instructed to write down each item as many times as possible during the rehearsal interval (3, 9, or 15 sec) after its presentation. While the number of rehearsals was the same for groups informed of the test of recall before and after the list's presentation, recall was superior for the group that knew of the test beforehand, and the difference remained constant over rehearsal intervals; however, recall scores for both groups increased up to the 9-sec rehearsal interval and remained constant after that. Although the authors offered little explanation for the effect of knowing about the test before the list's presentation, they did suggest that some process other than rote rehearsal was responsible. However, their conclusion was based in part on the assumption that the number of rehearsals used by subjects in both groups did not differ, and it is questionable whether their procedure was sufficient to properly assess all rehearsal. It is quite possible that subjects told of the test before the list's presentation repeated items covertly as they performed the relatively slow (1.7 sec per item) writing of the items.

Among other improvements, the experiment reported here offers a more viable control for covert rehearsal. Instead of written rehearsal, rapid oral rehearsal was used, the subjects being required to say each item out loud at a rate of three repetitions every 2 sec during the rehearsal interval (0, 2, 4, or 8 sec). In addition, the role of instructions was further examined by attempting to determine the time (in relation to presentation and rehearsal) at which awareness of a subsequent test of

recall is necessary in order for it to facilitate that recall. To this end, subjects were told which specific items were to be remembered (in addition to being rehearsed) either before the individual item's presentation (prior instructions) or at the end of the rehearsal interval (postrehearsal instructions).

Furthermore, the relationship of rote rehearsal and awareness was examined. Better recall of items the subjects were instructed to remember by prior instructions would indicate that encoding takes place either during rehearsal or during an item's presentation. The effects of rehearsal intervals would indicate the point in the encoding sequence at which awareness of a final recall is necessary and would clarify the role of rehearsal. Superior recall with prior instructions at all rehearsal intervals would indicate that awareness is important during an item's *presentation* and that long-term storage is independent of rote rehearsal; on the other hand, if superior recall with prior instructions increased as a function of rehearsal time, it would show that awareness during *rehearsal* is important. Similar performance with both prior and postrehearsal instructions would lead to one of two conclusions. If performance on items to be remembered was better than that on items to be rehearsed with both prior instructions and postrehearsal instructions, then awareness of a subsequent test of recall may facilitate long-term retention at any time, even after rote rehearsal has been completed. If recall improved with increased rehearsal with both prior and postrehearsal instructions, then the effect of awareness could be attributed entirely to rote rehearsal.

METHOD

Subjects

The subjects were 40 student volunteers from introductory psychology classes at the University of Notre Dame. Each student received a slight addition to his final course grade for participating in the experiment.

Stimuli

Stimulus items were 40 CCC trigrams (29% association value) drawn from Witmer (1935). Another 10 items of the same association value served as a practice list for free recall.

Design

A one between-subjects, two within-subjects design was used. Time of awareness was varied between subjects. The 20 subjects given *prior* instructions were told which of the 40 items to remember before the presentation of each item. The remaining 20 subjects were given *postrehearsal* instructions and were not told whether or not an item was to be remembered until after its rehearsal. Item type (*rehearse* or *rehearse and remember*) and length of the four rehearsal in-

tervals (*0, 2, 4,* or *8* sec) were manipulated within subjects. Each subject was presented 20 items to rehearse and remember and 20 items to rehearse, 5 at each combination of item type and rehearsal interval. The order of the conditions (item type × rehearsal interval) was randomized and was reversed for half of the subjects. The design was completely counterbalanced so that each stimulus item appeared in each condition an equal number of times across subjects.

Procedure

The practice list of 10 items was recalled before the experimental task. Each subject was run individually. Stimuli and instructions were presented via a PDP8/I computer and an oscilloscope display for 2 and 4 sec respectively. Each sequence was the same: a set of instructions was presented on the screen, followed by a trigram. Termination of the trigram on the screen was the signal for the subject to begin rehearsing it aloud at a rate of three repetitions every 2 sec, paced by the clicks of an electronic metronome (as the rehearsal interval permitted). In each repetition, the trigram was to be spelled aloud. The appearance of the next set of instructions served as signal to stop rehearsing the last item and prepare for presentation of the next item.

All subjects were told before the experiment that they would be expected to remember some of the items for later recall. For subjects given prior instructions, the instructions to "rehearse and remember the following item" or to "rehearse the following item" appeared before the presentation of each item. For subjects given postrehearsal instructions, the instructions to "remember the last item and rehearse the next" or to "rehearse the following item" appeared after the rehearsal of each item.

After all items had been presented, written free recall of the items to be both rehearsed and remembered and those to be rehearsed only was taken for 5 min. The importance of recalling all items remembered, not just those they had been asked to rehearse and remember, was stressed to all subjects.

RESULTS

A correct response was scored when all three letters of a trigram were recalled in the correct sequence. The task was difficult. The highest mean number of items correctly recalled under any of the 16 treatment combinations (time of awareness × item type × rehearsal interval) was .95 (*SE*, .23). Table 1 shows the mean number of items correctly recalled, and Figure 1 shows the mean proportion of items correctly recalled. The number of correct responses was used as the basic data in a one between-subjects, two within-subjects analysis of variance. There was no significant main effect attributable to rehearsal interval, that is, 0 or 2 or 4 or 8 sec. There were, however, significant main effects due to time of awareness, that is, prior or postrehearsal instructions [$F(1, 38) = 5.085$, $p < .05$], and item type, that is, items to be rehearsed and remembered or just rehearsed [$F(1, 38) = 11.63$, $p < .01$]. The interaction between item type and time of awareness was also significant [$F(1, 38) = 8.67$, $p < .01$].

Table 1. Mean number of correct responses (and *SE*) as a function of time of awareness, item type, and rehearsal interval

Group	Rehearsal interval (sec) 0	2	4	8
Rehearse and remember				
Prior instructions	.85 (.19)	.90 (.26)	.95 (.23)	.95 (.19)
Postrehearsal instructions	.45 (.15)	.55 (.13)	.40 (.15)	.45 (.11)
Rehearse				
Prior instructions	.40 (.16)	.45 (.17)	.45 (.13)	30 (.13)
Postrehearsal instructions	.45 (.16)	.55 (.13)	.40 (.15)	.30 (.11)

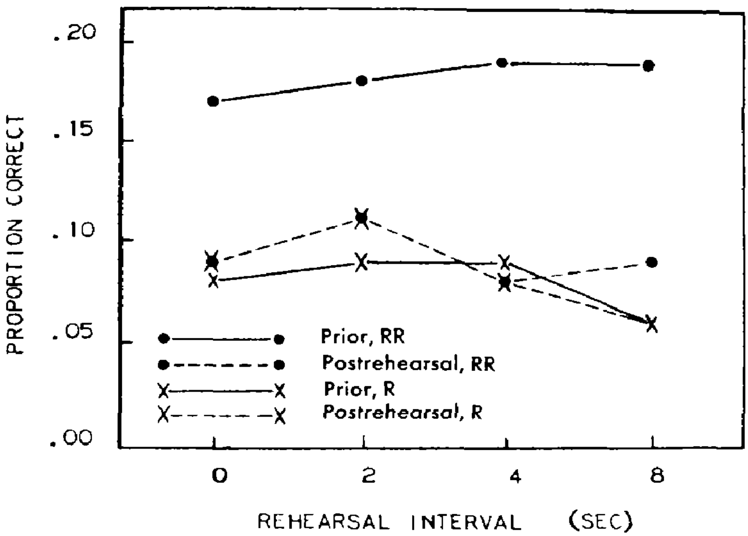


Figure 1. Proportion of items correctly recalled as a function of time of awareness (prior or postrehearsal instructions), item type (rehearse, R, or rehearse and remember, RR), and rehearsal interval

Since recall was quite low, revealing that a floor effect might be present, it was decided that a second, identical analysis using a less stringent scoring should be done. The number of items for which the first *two* letters were correctly recalled in the proper sequence served as the basic data for this analysis, and it showed exactly the same pattern as the first analysis. It was concluded that no floor effect was present.

A serial-position analysis is shown in Table 2, in which the mean number of correctly recalled items from each fourth of the presentation list

Table 2. Mean number of correct responses (and *SE*) as a function of list position, time of awareness, and item type

Group	List quarter			
	1	2	3	4
Rehearse and remember				
Prior instructions	1 15 (23)	1 30 (26)	65 (15)	.65 (.23)
Postrehearsal instructions	55 (.12)	.50 (.11)	25 (.10)	.55 (16)
Rehearse				
Prior instructions	55 (15)	.45 (.16)	35 (.13)	.25 (.12)
Postrehearsal instructions	.65 (13)	15 (08)	35 (.15)	.55 (.15)

is collapsed over rehearsal intervals. The superiority of the prior instructions to rehearse and remember an item is evident throughout the list.

DISCUSSION

The results have direct implications for the two issues of concern in this experiment: (a) the relationship between rote rehearsal and the effects of awareness of a subsequent test of recall and (b) the point at which such awareness is necessary in order for it to facilitate long-term retention.

The results make it possible to pinpoint some relationships between rote rehearsal and awareness that recall will be required. Some facts are certain. First, only with prior instructions (before presentation of the item) was there improved recall. Second, there was no increase in recall under any condition as a function of the amount of rehearsal, including the conditions with prior instructions. These facts lead to the following inescapable conclusions. First, rehearsal itself was of no benefit under any of the conditions; it is thereby ineffective in long-term storage and retention, at least as required by this task. Second, it is quite likely that the requirement to rehearse is *not* sufficient to ensure that the item is stored at all. That this may be the case is shown by the complete lack of a recency effect with prior instructions (see Table 2). Apparently, rehearsal can occur without 'storage in short-term memory.' Third, since rehearsal did not interact with any of the conditions, the number of rehearsals has no effect on the process responsible for long-term retention. Something else, then, must be responsible for long-term storage, and whatever the process is, it must be independent of rote rehearsal.

The results also indicate where in presentation and rehearsal that process takes place. As Figure 1 illustrates, better recall of items to be re-

hearsed and remembered resulted only when instructions were given *before* the presentation of an item. Learning that an item was to be remembered after it had been rehearsed had little effect on long-term retention. Evidently subjects use some type of long-term storage process during an item's presentation rather than its rehearsal, because the improvement attributable to prior instructions was present even when no rehearsal was permitted (the 0-sec rehearsal interval). Since there was no time for overt rehearsal in this case, the difference between prior instructions and postrehearsal instructions is entirely a matter of whether the subject was informed that an item was to be remembered before it was presented.

It is interesting to note that although subjects given postrehearsal instructions had ample time to store an item during the 4-sec presentation of instructions for the next item, they did not seem to do so. Apparently, they could not simultaneously place the item in long-term storage and process the instructions for the next item.

The results of the current experiment essentially replicate those of Meunier et al. (1974), and the collective results pose difficulties for some theories of learning. Both experiments demonstrate that something other than rote rehearsal is responsible for the effects of time of awareness, and both have shown that differences in retention between subjects with prior and postrehearsal instructions remain constant as the number of overt rehearsals increases. The present study adds credibility to the results of Meunier et al. (1974) by minimizing the possibility of covert rehearsal and by demonstrating the validity of the effect for the 0-sec rehearsal interval.

On a more general level, it appears that several of the current theories of memory (Waugh and Norman, 1965; Sperling, 1967; Atkinson and Shiffrin, 1968) are inadequate as explanations of the current data. They suggest that the likelihood of long-term storage may partially depend on the amount of rehearsal an item receives or its duration in temporary storage. It is obvious that this is not always the case. Subjects given prior and postrehearsal instructions repeated items an equal number of times, yet retention of the items they were instructed to remember was superior for the subjects given prior instructions. Thus, only when a subject knows before an item's presentation that recall will be required does awareness affect recall. The above theories fail to account adequately for this finding. While they propose that improved long-term retention may be due to the use of different strategies (Norman and Rummelhart, 1970) or 'control processes' (Atkinson and Shiffrin, 1967), they do not account for the fact that rote rehearsal has *no effect* on these processes.

Some clarification of the current findings is possible by examining the processing model proposed by Craik and Lockhart (1972). They suggested that retention depends on the level of analysis of the stimulus during encoding and that factors such as attention to the stimulus and the amount of time for processing are critical to long-term retention. They differentiate two levels of processing. Which level is used in learning is a function of attention, and the amount of long-term retention depends on the amount of time the deeper level is used. The current data suggest that subjects given prior instructions used deeper 'Type II' processing for items they were instructed to remember. Consistent with Craik and Lockhart's interpretation, 'Type I' processing (rote verbalization) did little to enhance retention. If subjects given prior instructions actually used Type II processing, then that processing may not include rehearsal, for in no case did rehearsal lead to better recall; no improvement in recall attributable to rehearsal interval was found. Indeed, the difference between prior and postrehearsal instructions at the 0-sec rehearsal interval further suggests that Type II processing is most effectively employed during an item's presentation.

Notes

Requests for offprints should be sent to Donald A. Walter, who is now at the West Virginia College of Graduate Studies, Institute, West Virginia 25112. Received for publication October 2, 1975; revision, February 9, 1976.

References

Atkinson, R. C., and Shiffrin, R. M. 1968. Human memory: A proposed system and its processes. In *Advances in the psychology of learning and motivation research and theory,* vol. 2, ed. K. W. Spence and J. T. Spence. New York: Academic Press.

Butterfield, E. C., Peltzman, D. J., and Belmont, J. M. 1971. The effects of practice upon rehearsal in short-term memory. *Psychonomic Science* 23:275–276.

Craik, F. I. M., and Lockhart, R. S. 1972. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior* 11:671–684.

Glanzer, M., and Meinzer, A. 1967. The effects of intralist activity on free recall. *Journal of Verbal Learning and Verbal Behavior* 6:928–935.

Jacoby, L. L., and Bartz, W. H. 1972. Rehearsal and transfer to LTM. *Journal of Verbal Learning and Verbal Behavior* 11:561–565.

Meunier, G. F., Ritz, D., and Meunier, J. A. 1972. Rehearsal of individual items in short-term memory. *Journal of Experimental Psychology* 95:465–467.

Meunier, G. F., Kestner, J., Meunier, J. A., and Ritz, D. 1974. Overt rehearsal and long-term retention. *Journal of Experimental Psychology* 102:913–914.

Norman, D. A., and Rummelhart, D. E. 1970. A system for perception and memory. In *Models of human memory,* ed. D. A. Norman. New York: Academic Press.

Restle, F. 1970. Training of short-term memory. *Journal of Experimental Psychology* 83:224–226.

Sperling, G. 1967. Successive approximations to a model for short-term memory. In *Attention and performance,* ed. A. F. Sanders. Amsterdam: North Holland Publishing. Reprinted from a special edition of *Acta Psychologica* 27(4).

Tulving, E. 1966. Subjective organization and effects of repetition in multitrial free-recall learning. *Journal of Verbal Learning and Verbal Behavior* 5(2)193–197.

Waugh, N. C., and Norman, D. A. 1965. Primary memory. *Psychological Review* 72:89–104.

Witmer, L. R. 1935. The association value of three-place consonant syllables. *Journal of Genetic Psychology* 47:337–360.

Scheduled for publication

Allen R. Dobbs. 'Forced' encoding but 'no' learning
William A. Ciszcwski and Charles F. Flaherty. Failure of a reinstatement treatment to influence negative constrast
James H. Neely. The effects of visual and verbal satiation in a lexical decision task
G. J. Syme and Lesley A. Syme. Spontaneous alternation in mice: A test of the mere-exposure theory
William C. Gordon. Similarities between recently acquired and reactivated memories in the production of interference
Pamela R. Sneed, David C. Brunts, and John H. Mueller. Instructions and grouped presentation in free recall
Joseph A. Sgro. The use of relief and nonrelief from shock in the double alleyway
W. B. Pavlik and Alexis C. Collier. Resistance to extinction of rats as a joint function of magnitude and schedule of reinforcement: A within-subjects analysis
Chok C. Hiew. Sequence effects in rule learning and conceptual generalization
Drake R. Bradley and Heywood M. Petry. Organizational determinants of subjective contours: The subjective Necker cube
Ruth H. Maki. Recall as a function of the relationship between cues and to-be-remembered words
Ryan D. Tweney and Dan Swart. Experimental control of reaction times to negative sentences
Leon Manelis. Frequency and meaningfulness of tachistoscopic word perception
Joseph J. Franchina, James R. Weeks, Jr., and Nanette Pais. Effects of situational similarity and goal-box placements on extinction after early nonreward
Joan Regan. Strategies in learning for recall and recognition tests
Thomas R. Herzog and Daniel J. Weintraub. Preserving the kinesthetic aftereffect: Alternating inducing blocks day by day
James W. Pellegrino and Judith Petrich. List-discrimination processes during transfer in free recall
Robert Sommer. A note on 'graphonyms'
H. D. Kimmel. Notes from "Pavlov's Wednesdays": Sensory preconditioning
James W. Aldridge and Michael T. Farrell. Long-term recency effects in delayed recall
Louis G. Lippman and Richard W. Thompson. The effect of shock on the exploratory behavior of rats in a complex maze
Charles F. Hinderliter and David C. Riccio. Long-term effects of prior experience in attenuating retrograde amnesia

continued on page 54

AMERICAN JOURNAL OF PSYCHOLOGY
March 1977, Vol. 90, No. 1, pp. 39–44

Segregation strategies in memory scanning

John C. Schmitt and C. James Scheirer
State University of New York at Binghamton

Mixed lists of digit pairs and consonant pairs were presented under Sternberg's scanning paradigm. The results indicated that the subjects partitioned this material into discriminable subsets but apparently only used that structure to reduce response latency when the test probes were negative. Probe familiarity is discussed as a basis for this strategy.

As Okada and Burrows (1973) have pointed out, it is essential for efficient retrieval that a subject be able to partition material in memory; otherwise, he would face the task of searching through all of memory for each retrieval. This observation is particularly relevant under Sternberg's scanning paradigm, where much evidence suggests that subjects exhaustively search all items in short lists when responding to a recognition probe as rapidly as possible. Sternberg (1966, 1969) has argued that this is because high-speed scanning is more rapidly executed, and therefore more efficient, than checking the list for a match after each item. In more complex retrieval situations, however, subjects have been shown to utilize more complex strategies, including an apparently self-terminating search when scanning for serial position (Sternberg, 1969, Experiment 7).

Other evidence has indicated that when memory loads are structured such that processing time could logically be reduced by searching through only subsets in memory, subjects do in fact utilize that structure. Naus, Glucksberg, and Ornstein (1972), Naus (1974), and Okada and Burrows (1973), for example, found that slopes of scanning functions (increased reaction time with each additional item in the memory load) were reduced with lists of categorized words. Williams (1971) found that sets of letters learned on different colored backgrounds produced search restricted to the appropriate colors.

The purpose of the research presented here was to further investigate the conditions under which subjects can partition material and use that segregation under Sternberg's scanning paradigm. The lists were composed

ubsets of letter pairs and number pairs, and the conditions were struc-
ed such that selective search of the letter pairs could be detected. In
lition, for some subjects recognition was immediate, while for others it
delayed or filled with a distractor task. Orthogonally, subjects were
ıer instructed to segregate (partition) the materials or not so in-
ıcted. These retentional and instructional variables were manipulated
ın attempt to allow and induce segregation.

THOD

paratus and materials

ists of one, two, or three number pairs and/or letter pairs and single probe rs were mounted on 35-mm slides and projected by a Carousel projector. it pairs were constructed by randomly sampling without replacement from digits 2–9, and consonant pairs by selecting bigrams from the pool of English sonants P–W, again without replacement. Hunter timers controlled all timing, uding that of list presentation and offset, retention interval, probe presenta-ı, and the offset and onset of an Eldorado digital timer (model 1410) that ısured reaction time. A response panel contained two buttons, labeled 'yes' 'no,' and was used by the subjects to make their responses. Pulses from these ons stopped the Eldorado timer and began the intertrial interval. The posi-s of the buttons were balanced across subjects in order to control for dedness.

cedure

or *condition I,* the lists consisted of one, two, or three letter pairs, where a d of the lists were of each of the three lengths. This replicates the standard ning procedure. For *condition II,* the lists always contained one letter pair ther with no, one, or two number pairs, such that the lists were of each th (one, two, or three pairs) on a third of the trials. For *condition III,* the contained one, two, or three letter pairs, each on a third of the trials, in unction with two, one, or no number pairs, such that all lists contained three s total. The purpose of these conditions was to vary independently the num-of pairs in a list and the number of letter pairs, in order to determine ther subjects search all pairs in the memory load or only the relevant (let-ones.

ıch trial consisted of the presentation of a list of items followed by a short ıtion interval, which was followed by a 5-sec test probe. In all cases, the (s) making up a list appeared on a single slide for a duration of 1 sec per After the retention interval, a single probe pair was presented; it either (positive) or was not (negative) a member of the list. The probes were nced for positive and negative occurrences and for letters and numbers on rials where numbers were list items. Where no numbers were present on ist, only 30% of probes were numbers. Trials were blocked in 10 replica-of the six combinations of probe type and list length, and preceded by 12 tice trials for a session of about 40 min.

addition to the manipulation of conditions (lists), three retention intervals

and two types of instructions were employed. *Immediate* retention consisted of probe presentation approximately 1 sec after list offset. *Delayed* retention with a *distractor task* consisted of presentation of a four-digit number 1 sec after the list's offset and 7 sec of counting backward by threes (see Peterson and Peterson, 1959). *Delayed* retention with a *blank interval* consisted of 8 sec between list and probe, with no distractor presented.

Segregation instructions added to the standard instructions the statement, "Previous research has indicated that segregation of the list items into separate number and letter sets aids performance on this task." Half of the subjects in condition I were instructed to "unify" the set. *Standard* instructions were simply to respond as rapidly as possible consistent with errorless performance. The design in entirety thus yielded 18 independent groups of subjects (three conditions × three retention intervals × two instructions).

Subjects

The subjects were 108 students in introductory psychology at the State University of New York at Binghamton. They participated in order to satisfy a course requirement. Subjects were randomly assigned to conditions in blocks of the 18 experimental groups.

RESULTS AND DISCUSSION

Harmonic mean latency of correct response to letter probes was the dependent variable. Error trials accounted for less than 3% of all trials, were not systematic with either within-subject or between-subject manipulations, and were discarded. Overall, responding was faster to positive probes than to negative ones, 994 versus 1,058 msec [$F(1, 90) = 23.67$, $p < .01$]. Because of differential results for positive and negative probes, they were analyzed separately.

Positive probes

While immediate testing produced faster responding and a distractor task further slowed responding [all ps $< .01$], retention interval did not interact with any other manipulations [all ps $> .05$]. Segregation instructions had no overall effect on response latencies [$F < 1$], nor did they interact with any other manipulations [all ps $> .05$]. It should be noted that interactions with slope are analyzed within subjects; thus the nonsignificance obtained does not reflect a lack of power in the analysis.

Conditions I and II did not differ in their effect on overall responding to positive probes [$F < 1$], but they did, on the average, produce faster responding than did condition III, 973 versus 1,132 msec [$F(1, 90) = 6.08$, $p < .05$]. Furthermore, the slopes of the scanning functions in conditions I and II showed highly significant linear increases [$F(1, 450) =$

Accession numbers

71.97 and 63.43 respectively, *p*s $< .01$], while that of condition III was flat [$F(1, 450) = 1.42, p > .05$]. The slopes of 147, 137, and −21 msec per item for conditions I, II, and III respectively showed no difference between conditions I and II [$F < 1$], but the average of these was significantly greater than that of condition III [$F(1, 450) = 59.10, p < .01$]. These relationships are illustrated in the left panel of Figure 1. They are, of course, exactly what would be expected if subjects were 'pooling' the items, that is, using no apparent segregation of the lists' structure in responding. These reaction times reflect number of pairs, rather than number of letter pairs.

Negative probes

As was the case with positive probes, responding to negative probes was slower at delayed retention than at immediate testing and slower with a distractor task than without [*p*s $< .01$], but retention interval did not interact with any other manipulations [all *p*s $> .05$]. Segregation instructions were not a significant source of variance [$F(1, 90) = 2.32, p > .05$] and did not interact with any other manipulations [all *p*s $> .05$].

Levels of responding to negative probes between conditions I and II showed no differences [$F < 1$], but the average of these conditions did produce faster responding than did condition III, 1,008 versus 1,160 msec [$F(1, 90) = 5.13, p < .05$]. The scanning functions for conditions I, II, and III, shown in the right panel of Figure 1, all showed highly significant linear trend components [$F(1, 450) = 126.54$, 15.80, and 72.56 respectively, $p < .01$]. While the slopes of conditions I (195 msec per item) and III (147 msec per item) did not differ [$F(1,450) = 3.77, p > .05$], their average trend was greater than that of condition II (69 msec per item) [$F(1, 450) = 10.29, p < .01$]. This result, then, is consistent with a segregation of memory sets, since the reduced slope in condition II indicates a tendency to respond only to the number of letter pairs.[1]

Any effect of number of letter pairs on slopes that is independent of total pairs suggests a segregation of material under this paradigm, and perhaps in immediate memory in general. This idea is consistent with the result that segregation instructions had no effect with a distractor task. It would therefore appear that at some stage in processing subjects partition material into discriminable subsets. The question is why this information was only used in the case of negative probes, which were unpredictable as to occurrence.

One idea receiving some support from these data utilizes notions similar to the response criteria of Atkinson and Juola (1974). They proposed that

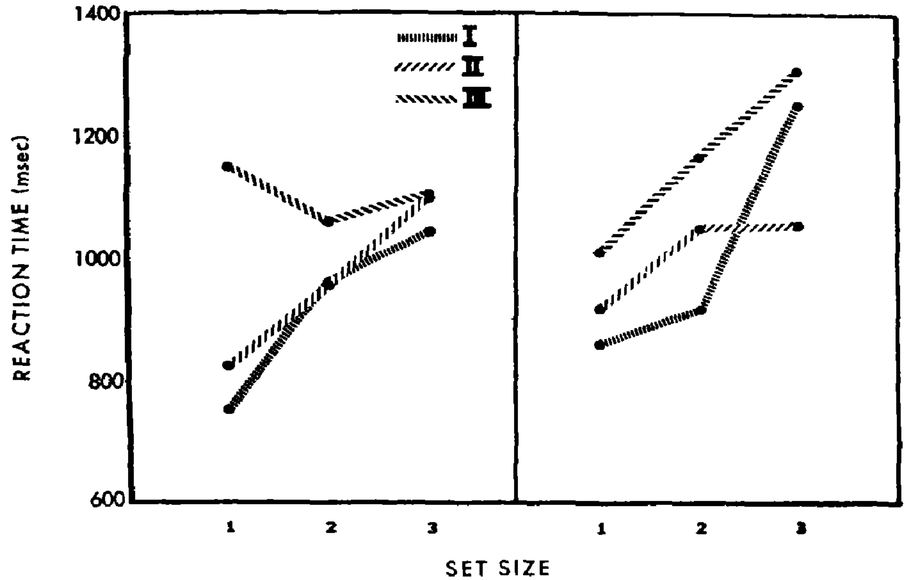


Figure 1. Scanning functions for positive (left panel) and negative (right panel) probes for each condition; set size refers to total number of pairs presented under conditions I and II, but to the number of letter pairs presented under condition III

the probe's place on a continuum of familiarity at the time of testing is used by the subject to determine his response strategy. With the probe's familiarity above or below two criteria, the subject immediately responds positively or negatively, with no search, while at intermediate familiarity, scanning occurs. Our data indicated segregated search when the test probe was negative but not when it was positive. The strategy here may therefore be due to the negative probe's falling low enough on the continuum that a search of only the appropriate type of material is made, though not low enough to produce an immediate negative response. Once no match is found, the negative response is made with no further search of the irrelevant set. In this case, the subject would be scanning in only a small proportion of the test trials, thus reducing the apparent scanning slope for negative probes (condition II). Positive probes, however, could be expected to fall somewhat higher on the continuum, above some criterion, and in this situation, a search is made of the complete memory set without regard to type of material.

While this explanation is consistent with the results and not implausible, why would subjects abandon the segregated structure in the case of positive probes? It is possible that for a negative probe below the criterion,

the search is made only to reaffirm the initial evaluation that it was not a member of the list; for a positive probe, the initial decision is that it *was* on the list and a scan of the entire list is made so that the subject may be sure to locate it. In any event, these findings suggest that a specific strategy may occupy a specific band along an internal continuum of the familiarity of the probe. As Reitman (1970) has suggested, the analysis of these strategy systems "reflects the gradual alteration in our notions of reasonable experimental procedure" (p. 504) and is an important future direction for research on information processing.

Notes

The authors thank Michael Hanley, William Lutz, and Richard Friedlander for their help with various stages of this research. Requests for offprints should be sent to C. James Scheirer, Department of Psychology, State University of New York at Binghamton, Binghamton, New York 13901. Received for publication April 9, 1976.

1. An earlier study using immediate retention and standard instructions, but different equipment, perfectly replicated the study reported here. Details and results of that study are available from the second author.

References

Atkinson, R. C., and Juola, J. F. 1974. Search and decision processes in recognition memory. In *Contemporary developments in mathematical psychology,* ed. D. H. Krantz, R. C. Atkinson, R. D. Luce, and P. Suppes. San Francisco: W. H. Freeman.

Naus, M. J. 1974. Memory search of categorized lists: A consideration of alternative self-terminating search strategies. *Journal of Experimental Psychology* 102:992–1000.

Naus, M. J., Glucksberg, S., and Ornstein, P. A. 1972. Taxonomic word categories and memory search. *Cognitive Psychology* 3:643–654.

Okada, R., and Burrows, D. 1973. Organizational factors in high-speed scanning. *Journal of Experimental Psychology* 101:77–81.

Peterson, L. R., and Peterson, J. J. 1959. Short-term retention of individual verbal items. *Journal of Experimental Psychology* 58:193–198.

Reitman, W. 1970. What does it take to remember? In *Models of human memory,* ed. D. A. Norman. New York: Academic Press.

Sternberg, S. 1966. High-speed scanning in human memory. *Science* 153:652–659.

Sternberg, S. 1969. Memory scanning: Mental processes revealed by reaction-time experiments. *American Scientist* 57:421–457.

Williams, J. D. 1971. Memory-ensemble selection in human information processing. *Journal of Experimental Psychology* 88:231–238.

AMERICAN JOURNAL OF PSYCHOLOGY
March 1977, Vol. 90, No. 1, pp. 45–53

A comparison of item and position probes in short-term memory

Douglas K. Detterman
Case Western Reserve University

The tasks with item and position probes seem similar. Given an item probe, a subject must recall its position in the spatial array; given a position probe, the item in that position in the array. The two tasks should yield equivalent results only if item and order information are dependent and therefore subject to the same processes. Half of the 46 subjects were tested with item probes, half with position probes, in otherwise identical tasks. Analysis of correct responses and latencies showed that item and position probes yielded different results. These differences are interpreted as reflecting the independence of item and order information.

It is often implicitly assumed that all tasks measuring short-term memory measure the same processes even though the tasks are operationally quite different in what they require of the subject. Two tasks that appear to be very similar — and are therefore seldom differentiated according to the memory processes they are supposed to measure — are those with item and position probes. In both tasks, a series of items is presented within a spatial array. The tasks differ only with respect to the type of probe presented. Given an *item* probe — one of the stimulus items of the original series — the subject is to recall the position in which that item occurred. Given a *position* probe — one of the serial positions of the list — the subject is to recall the item which appeared in that position. Since the nominal stimulus sets for both types of probe are identical, it is tempting to assume that both measure the same processes. Is this actually the case?

The answer to this question rests on the relationship between order and item information in short-term memory. If the two types of information are perfectly dependent, both types of probe should yield the same results. On the other hand, if different processes are operating in the retention of each type of information, retention as measured by an item probe should differ from retention as measured by a position probe. The pur-

pose of the present experiment was to determine if the tasks with the two types of probe produce similar or different results.

A number of experiments have reported comparisons between tasks using various types of probes (e.g., Murdock, 1968; Woodward and Murdock, 1968). However, the purpose of these experiments was not to determine the relationship between item and order information, and so the kinds of probes selected were not appropriate for such a determination. Probably the most frequent comparison made has been between sequential and position probes. In a task with a sequential probe, the subject is given item $n - 1$ and required to recall item n. In a task with a position probe, the subject is given the position and required to recall the item which occurred in that position. Thus, in both kinds of task, subjects are required to recall item information. These experiments were generally considered to be tests of the chaining and ordinal-position hypotheses of serial learning.

Even though there have been no explicit comparisons of tasks with item and position probes, there is good reason to think that the two tasks might produce different results. A number of experiments have found that item and order information appear to be retained independently. The experimental evidence leading to this conclusion has been discussed by Detterman and Brown (1974) and will not be repeated here. Since then, an experiment by Bjork and Healy (1974) has come to a similar conclusion. In all of the published experiments to date, the independence of item and order information has generally been established by showing that different variables have different effects on the two kinds of information or that the same variable influences retention of the two kinds of information differently. The present experiment seems to be a clearer test of the independence of item and order information, since the subjects' performance is more directly dependent on one type of information.

METHOD

Subjects

The subjects were 46 students enrolled in introductory psychology courses. They received extra course credit for participation and were assigned alternately to conditions as they reported.

Apparatus

The apparatus consisted of an IBM memory drum (Lafayette 23017) modified for presentation of the task. The 13.5-in. slit on the face of the memory drum was covered with a cardboard mask having eleven openings, each .75 in. square, spaced 1.25 in. from center to center. Each of the first ten openings from the left was labeled beneath with a number indicating its serial position. The last opening was labeled 'P,' designating the point at which the probe was displayed.

The memory drum accepted IBM computer paper, and all stimulus materials were produced on a computer printer. Sequential presentation of stimulus items in each of the openings of the mask was accomplished by printing each list on a diagonal running from the upper left to the lower right corner of the page. Each advance of the memory drum displayed a single item in successive openings proceeding from the left to the right of the mask.

Stimuli

Stimulus items for each list were the first ten letters of the alphabet (A–J) presented in a random arrangement. This stimulus set has undesirable features (e.g., a high degree of acoustic confusability) but was chosen because it appeared to be as well defined a set as the ordinal numbers (1–10) that represented serial positions. There were 52 lists in all. The first 2 were practice lists, and the rest made up the experimental task. Arrangements of the stimulus letters of each of the 50 lists of the experimental task were randomly selected with the exception that each letter appear at each serial position equally often across lists. The probe for each list was also randomly selected with the restrictions that within each block of 10 lists, each item and each serial position be probed equally often.

Procedure

The type of probe that was presented differentiated the two experimental conditions. In the *position-probe condition,* each list was followed by one of the arabic numbers 1–10 and the subject was required to recall the item which had occurred in that serial position. In the *item-probe condition,* each list was followed by one of the letters A–J and the subject was required to recall the serial position in which that item had occurred. For any given list, the probe in both conditions tapped the same item in the same serial position. The only thing that differed between conditions was the type of information requested of the subject. The item-probe condition required recall of order information and the position-probe condition required recall of item information.

Subjects participated in the experiment individually. The instructions stressed the importance of accuracy and indicated that speed of response was of secondary concern. To begin each trial, the experimenter said "Ready" and then started the memory drum, which presented the stimuli at a rate of 2 sec per item. When all items on the list had been presented and the probe appeared, the experimenter stopped the memory drum. For both conditions, the latency of each response was measured by a timer calibrated in hundredths of a second. The timer was activated when the experimenter stopped the memory drum on the probe and was terminated by the subject's vocal response into the microphone of a voice key.

RESULTS

The primary data for analysis consisted of the number of correct responses made to each serial position. In addition, latencies to correct and incorrect responses were analyzed separately. During testing, each serial

position (or the item appearing in that serial position) was probed five times. For some subjects, all responses to a particular serial position were either all correct or all incorrect, which meant that no estimate of latencies to incorrect or correct responses respectively was available at that serial position. To compensate for this, latencies were combined for pairs of successive serial positions. Even when blocked in this fashion, the data for four subjects in each condition had to be eliminated, three in each condition for failure to make a correct response to either of two successive serial positions and one in each condition for failure to make an incorrect response to either of two successive serial positions. Fortunately, the reasons for eliminating these subjects' data were the same for both conditions, which suggests that the blocking had no differential effects on experimental conditions but rather was necessitated by similar effects of task length and difficulty on each condition. For convenience in interpretation, the analysis of correct responses was based on data from the same subjects as the analysis of latencies, though not blocked across serial position.

Correct responses

Figure 1 shows the mean percent correct responses, by serial position, under each condition. Under the position-probe condition, average correct recall of item information was 50.4%. Under the item-probe condition, average correct recall of order information was only 37.3%. Analysis of variance confirmed the superior retention evidenced under the position-probe condition. The main effect of condition was statistically significant [$F(1, 36) = 14.99, p < .001$]. In addition, the interaction of serial position by condition [$F(9, 324) = 3.35, p < .01$] indicated that the largest differences in retention between the two conditions occurred at the beginning and end of the list.

Incorrect responses

One way to determine if subjects under the two conditions were focusing on different aspects of the stimulus situation or demonstrating different response biases is to analyze incorrect responses according to each of the stimulus dimensions. For example, for the position information, a matrix is formed with each column representing the correct position of a particular response and each row representing the position of the response actually given. The principal diagonal represents correct responses. Successive diagonals parallel to the principal diagonal represent incorrect responses, successively further removed from the correct position. If a subject indi-

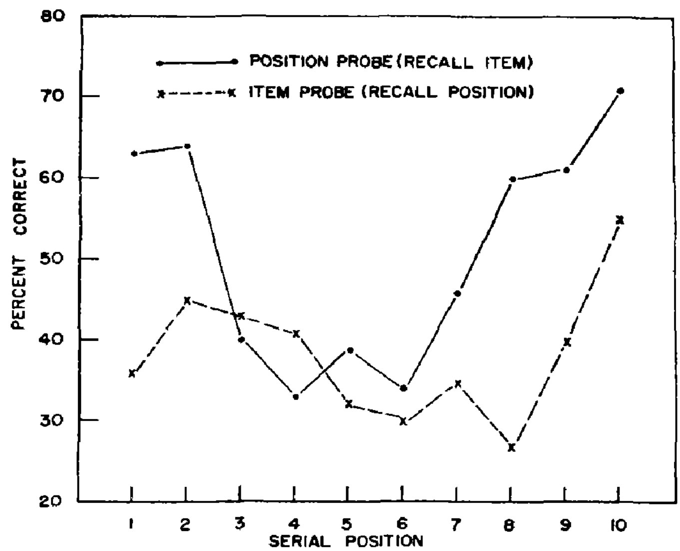


Figure 1. Percent correct responses by serial position for the position- and item-probe conditions

cated that a stimulus had occurred in the seventh position but it actually had occurred in the fifth position, then this incorrect response would be tallied in the cell represented by the intersection of the fifth column and seventh row, which is the second diagonal away from the principal diagonal. A similar analysis can be carried out on item information if it is assumed that letters of the alphabet form an ordered stimulus set. Detterman (1974) found that in a probe task for lifted weights, subjects organized both item and position information according to ordered dimensions, and these dimensions were independent. Thus, there is reason to believe that subjects' retention of item information is represented by an ordered dimension. What this means in the present case is that if the correct response was D, the subjects might be expected to show a tendency to select incorrect responses close to it, say, C or E, if they were using an ordered dimension of item information.

Note that the analysis of incorrect responses for each type of stimulus information is different for each condition of the experiment. As an example, take the analysis of position information. Under the item-probe condition, the subject gives a position as a response. Tallying this piece of data is straightforward; the correct response is the column and the response given is the row. Under the position-probe condition, the subject gives a letter as a response. To tally this piece of data in the matrix for

position information, the *probe* is the column and the position in which the letter given by the subject occurred is the row.

Once matrices have been formed for item and position information for each condition and incorrect responses tabulated, each diagonal away from the principal diagonal may be summed over cells. What results is what may be thought of as a 'generalization' gradient indicating how close to the correct response the incorrect responses were. Table 1 shows the results of this analysis. Incorrect responses *n* places before and after the correct response (i.e., diagonals *n* places above and below the principal diagonal) were added together, since they appeared similar. In Table 1, these sums are expressed as percentages of total incorrect responses for each condition. The distribution of incorrect responses that would be expected on the basis of chance is also shown. The χ^2 values compare each kind of stimulus information for each condition to the results expected by chance. These comparisons show that all are different from chance except for item information under the position-probe condition. The major point of this table, though, is that subjects under both conditions seemed to be responding in a very similar manner and to the same sources of information. There appears to be no strong evidence for differential response bias.

Latencies of correct and incorrect responses

Figure 2 shows mean latencies to correct and incorrect responses by blocks of two serial positions under each condition. Analysis of variance of mean latencies to incorrect responses indicated that only the main

Table 1. Percent incorrect responses made at each position away from correct response for item and order information

Probe type	Positions away from correct response									χ^2
	1	2	3	4	5	6	7	8	9	
Item information										
Item	19 2	21.1	19 2	12.6	10.2	7.8	5 1	2.7	2.2	16.39*
Position	19.3	19 3	17.4	13.3	12.1	7 2	6 2	4 5	.8	8 32
Position information										
Item	35.9	21.5	17 9	10.8	6 2	4 6	2 5	.3	.3	149.93*
Position	32.9	19 1	15.7	11 4	9.1	5.8	3 5	.8	1.6	71 40*
Expected by chance										
	20.0	17 8	15.6	13.3	11.1	8.9	6.7	4.4	2.2	

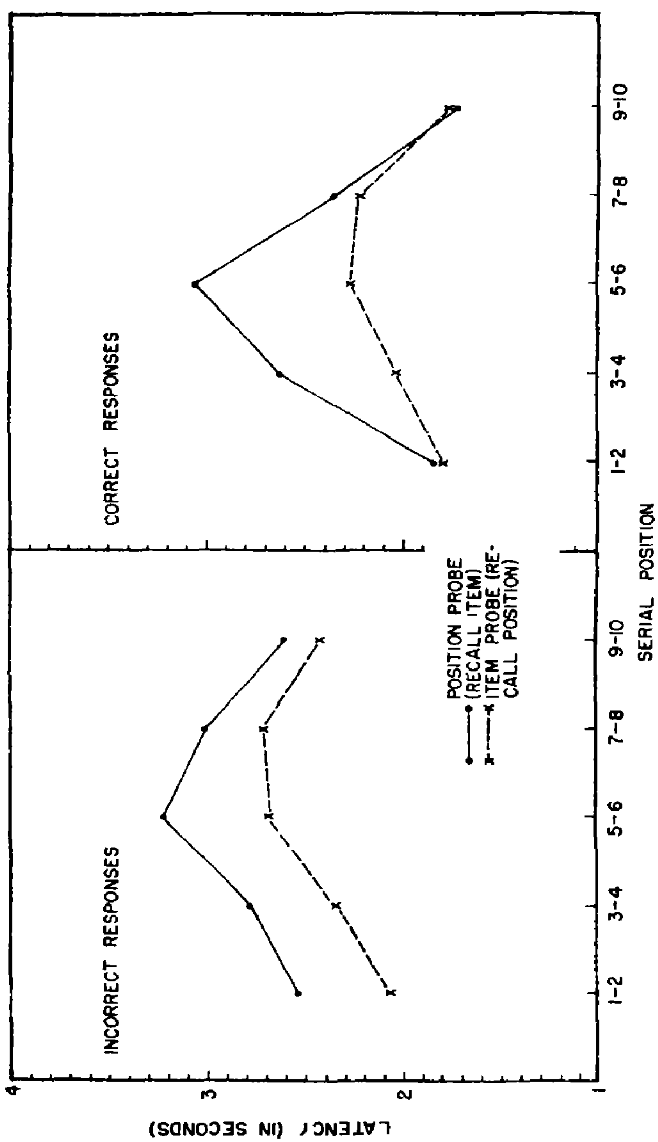


Figure 2. Latencies to correct and incorrect responses for the position- and item-probe conditions

effect of serial position [$F(4, 144) = 4.00, p < .01$] was significant. A similar analysis of mean latencies to correct responses revealed a significant main effect of serial position [$F(4, 144) = 10.86, p < .001$] and an interaction of serial position by condition [$F(4, 144) = 2.49, p < .05$]. The interaction reflected longer latencies for items in the middle of the list under the position-probe condition.

DISCUSSION

It is clear from consideration of the correct and incorrect responses and the latencies associated with those responses that the tasks with the two types of probe were different. Both dependent measures showed markedly different patterns of performance in the two tasks. Under the position-probe condition, in which subjects recalled item information, retention was markedly better for first and last items than for items in the middle of the list. Latencies to correct responses were shortest for first and last items and longest for items in the middle. On the other hand, under the item-probe condition, in which subjects recalled order information, retention was not as closely related to serial position. That is, the serial-position curve for that condition was less bowed. The same holds true for latencies to correct responses. Though there were differences in speed of responding across serial position, they were not as pronounced as under the position-probe condition.

The main effects between the item- and position-probe conditions are not interpretable as indications of the independence of item and order information. They might only reflect differences in difficulty between the two response sets. However, the same cannot be said of the interactions of condition by serial position for correct responses and for the latencies to these responses. It seems far more parsimonious to interpret these interactions as reflecting the independence of item and order information. This conclusion is consistent with a growing body of evidence that the two kinds of information are independent. The results of this study also suggest that the spatial-probe task may be a fruitful one for use in the investigation of the retention of item and order information.

Notes

Requests for offprints should be sent to Douglas K. Detterman, Department of Psychology, Case Western Reserve University, Cleveland, Ohio 44106. Received for publication April 9, 1976.

References

Bjork, E. L., and Healy, A. F. 1974. Short-term order and item retention. *Journal of Verbal Learning and Verbal Behavior* 13:80–97.

Detterman, D. K., and Brown, J. 1974. Order information in short-term memory. *Journal of Experimental Psychology* 103:740–750.

Detterman, D. K. 1974. Distinctiveness and serial position in short-term memory for lifted weights. *American Journal of Psychology* 87:95–106.

Murdock, B. B., Jr. 1968. Serial-order effects in short-term memory. *Journal of Experimental Psychology Monograph* 76(4, part 2):1–15.

Woodward, A. E., and Murdock, B. B., Jr. 1968. Positional and sequential probes in serial learning. *Canadian Journal of Psychology* 22:131–138.

continued from page 38

Henry M. Halff. The role of recall opportunities in learning to retrieve

John P. Forgas and L. B. Brown. Environmental and behavioral cues in the perception of social encounters: An exploratory study

Donald H. Kausler, David Kleim, and Robert Yadrick. False-recognition effect for synonyms of right and wrong items

Lois M. Freed and William J. Freed. Effects of schedules of reinforcement on producing hypothesis-refining behavior

Douglas L. Hintzman and Leonard D. Stern. Memory for items presented after massed and spaced repetitions: Failure to confirm Elmes, Greener, and Wilkinson's result

John C. Thomas, James L. Fozard, and Nancy C. Waugh. Age-related differences in naming latency

James D. Evans. On the inconstant effects of study instructions

John C. Schmitt and C. James Scheirer. Task difficulty as a mediator of subject strategy in memory scanning

John J. Shaughnessy. Long-term retention and the spacing effect in free recall and frequency judgments

Douglas J. Bobko, H. R. Schiffman, Robert J. Castino, and Wayne Chiappetta. Contextual effects in duration experience

Mukul K. Dey. Anagram-solution speed as a function of manifest anxiety and number of category sets

Fred Attneave and Paul Farrar. The visual world behind the head

Lawrence M. Ward, Clare Porac, Stanley Coren, and Joan S. Girgus. The case for misapplied constancy scaling: Depth associations elicited by elusion configurations

Christine D. Bremer, John B. Pittenger, Rik Warren, and James J. Jenkins. An illusion of auditory saltation similar to the cutaneous 'rabbit'

Lynn Hasher and Michael Greenberg. Expectancies as a determinant of interference phenomena

Donald A. Walter. Discrete events in word encoding: The locus of elaboration

John L. Locke. Large auditory and small visual effects in the recall of consonants

Paul T. P. Wong. Strategy learning and the skill partial-reinforcement effect in the rat

Seth N. Greenberg. Competition between memory and perceptual tasks involving physically similar stimulus material

Peter M. Lisiecki and Terry M. Libkuman. The relative effectiveness of rule 1 and rule 2 in verbal-discrimination learning

AMERICAN JOURNAL OF PSYCHOLOGY
March 1977, Vol. 90, No. 1, pp. 55–61

An independence of induced amnesia and emotional response

Dennis Runcie and R. Michael O'Bannon
University of Alabama

A photograph of a nude was interpolated midway through a 30-item list. Recognition memory of items at various serial positions was measured by presenting 12 old and 12 new pictures on a test trial. Palmar conductance was also measured. Significantly decreased recognition memory and increased palmar conductance accompanied presentation of the picture of the nude. When the two measures were compared for individual subjects, however, no correlation was found. These data suggest that both responses are likely to occur to the presentation of the critical item but that the responses are independent.

In the learning or remembering of otherwise homogenous lists, the presence of a unique item results in enhanced memory for that item. The improved memory for the unique stimulus, however, has often been at the expense of those items which surround it, either spatially or temporally. This effect is commonly known as the von Restorff phenomenon (see Wallace, 1965, for a review). It has been tempting to relate any memory deficit obtained under these conditions to a type of amnesia resulting from the presence of a unique, outstanding, or 'traumatic' item.

Tulving (1969) used a free-recall task to test memory for lists of serially presented common words. Some lists contained 'high-priority' words that were made unique by instructions. At relatively slow rates of presentation, he found some amnesia for words preceding the high-priority word, and he discussed this finding in terms of consolidation theory. Ellis, Detterman, Runcie, McCarver, and Craig (1971) got similar results when recall was used to measure memory for a 'list' of pictures which, on some trials, contained a nude picture. Their main finding, however, was based on recognition memory for 30-item lists composed of photographs from magazines, each list containing, at serial position 15, a photograph of a nude from a 'sunbathing' magazine. They observed that the deficits in recognition memory could be localized at the serial positions immediately fol-

lowing the nude picture, and they suggested that the presence of a unique item in a list taxes the capacity of an information-processing system to effectively handle subsequent information.

In ascribing to a unique stimulus certain 'amnesic' properties, the assumption is commonly made, either implicitly or explicitly, that it elicits an emotional response which in some manner interferes with the processing of adjacent stimuli (Tulving, 1969; Saufley and Winograd, 1970; Erdelyi and Appelbaum, 1973). This assumption has not been subjected to direct test. The unique stimuli chosen by Erdelyi and Blumenthal (1973) and Ellis et al. (1971) were undeniably emotional in nature ("a horribly deformed child's face" and nude photographs respectively), but indirect evidence cited by Ellis et al. (1971) suggests that the emotional quality of the stimuli may have been incidental to their amnesic effect. It would be expected that the emotional consequences of the critical stimulus would be felt for a period of time determined by the shock value of the item. Contrary to this expectation, Ellis et al. (1971) noted that deficits in recognition memory involved the same number of items whether the rate of presentation was .75 or 1.5 sec per item. Detterman and Ellis (1972) varied the rate of presentation of the neutral and critical items independently and found no effect attributable to the duration of the latter. In the present context, this would suggest that the item-dependent amnesic effect described by Ellis et al. (1971) was not due to changes in emotionality attributable to different exposure durations of the critical item.

The purpose of the present investigation was (a) to determine whether or not an emotional response, as measured by palmar conductance, does accompany a critical item, and (b) if it does, to investigate its relationship to the deficit in recognition memory.

METHOD

Subjects

The 40 male and female undergraduates who were recruited from an introductory psychology class received credit for serving as subjects. The subjects were randomly assigned to one of two groups of 20 each. Each subject was run individually.

Apparatus

Changes in palmar skin conductance were measured between two electrode cups filled with zinc electrode jelly. Electrode cups were located on the palm of the subject's right hand and were separated by about 5.0 cm. Resistance was measured on a Fels Dermohmmeter (model 22A) and recorded on an Esterline-Angus chart recorder. Stimuli were photographed on 16-mm film and presented

on a projector (L-W 800) modified to allow the presentation of a single frame for a duration specified by the experimenter. The subject was seated approximately 2 m from the projection screen. The film projector was located approximately 2 m behind, and in the same room as, the subject and projected an image approximately 1 m high. The experimenter controlled the projector and recorded the subject's 'yes' or 'no' responses. A voice-operated relay was used to signal response latency, which was timed on a Hunter clock-counter. A second experimenter, located in a separate room, read and recorded latencies to within 1 msec and monitored the Dermohmmeter.

The stimuli were a mixture of color and black-and-white photographs taken from popular magazines. Seventeen lists of these stimuli were used, with each of these *learning* lists containing 30 items drawn at random from a pool of 870 photographs and presented at the rate of 1 sec per frame. After each learning list was a *test* list of 12 randomly ordered stimuli from the preceding list and 12 the subjects had not previously seen. For the test list, the presentation rate was 3 sec per frame.

For subjects in the *experimental* group, a photograph of a nude was included in the learning list at serial position 15 on seven trials, the fourth, seventh, eighth, tenth, twelfth, sixteenth, and seventeenth. Subjects in the *control* group saw the same learning and test lists with the exception that the photograph of the nude was replaced with a neutral stimulus in each case. There was no duplication of stimuli in the learning lists, nor did a new test stimulus appear more than once. On a list containing a nude photograph, or the corresponding list for the control group, the serial positions probed were 3, 6, 9, 12, 13, 14, 16, 17, 18, 21, 24, and 27. On all other lists, the remaining serial positions were probed on a random basis, with the stipulation that each position be probed equally often.

Procedure

Each subject was invited into the experimental room, seated in a comfortable chair, and electrodes were affixed. A 15-min rest period then followed, during which the subject became habituated to the apparatus and a stable baseline level of palmar conductance was established. After this rest period, the subject was read instructions about the nature of the task and was told to respond yes to test pictures he had seen in the learning list and no to pictures he had not seen there. He was further instructed to respond rapidly once he had decided a picture was new or old. After a question period, the learning and test lists were presented. A 5-min rest period was given after the ninth trial to allow the subject to move about; the experimental session lasted about 50 min.

RESULTS

Scores included for analysis were the numbers of correct responses at each serial position sampled on the seven relevant trials (subjects' scores ranged from 0 through 7). The autonomic measure of emotional response was defined as the amount (mm) of the stylus' deflection at the presentation of each item in the learning list. The lineal measure converted to ohms provided the data base for all analysis. This was subsequently trans-

formed to conductance for graphic representation. Separate analyses were performed on items preceding (serial positions 1–14) and following (serial positions 16–30) the critical item.

Recognition memory

Figure 1 illustrates the percent correct identification of items from the learning lists on the seven trials at issue. A reduction in accuracy of recognition memory was clearly indicated both before [$F(1, 38) = 6.40, p < .05$] and after [$F(1, 38) = 28.53, p < .001$] the critical item. An inspection of the upper panel of Figure 1 suggests poorer recognition at serial positions immediately following the critical item. The interaction of serial position by treatment reveals that this was the case for the second half of the list [$F(5, 190) = 3.95, p < .01$]. Subsequent t tests indicated significant differences at serial positions 16, 17, 18, and 21. A significant interaction was also obtained for the first half of the list [$F(5, 190) = 3.58, p < .01$]. Subsequent t tests revealed significant differences at positions 9, 13, and 14, but not at serial position 12. The lack of significance at position 12 and the ambiguity of the deficit before the critical item do not allow clear-cut interpretation of this interaction.

The possible effect of repeated presentation of the nude photograph to the experimental group was assessed by comparing group differences at serial positions following the nude stimulus on the seven trials at issue. In addition to the main effect of groups, already shown to be significant, an improvement in performance over trials occurred for both groups [$F(6, 288) = 18.7, p < .01$]. No interaction of groups by trials was obtained [$F(6, 288) = .95$], indicating that the improvement in performance of the experimental group as the subjects became accustomed to the nude photograph did not exceed the improvement of the control group. An analysis of 'false alarms' by the two groups revealed no significant difference between their means, 1.96 for the experimental group and 1.52 for the control group.

Palmar conductance

The data on emotional response were treated like those on recognition memory, in that separate analyses were performed on items preceding and items following the critical stimulus on each of the seven trials. No difference between groups was found in either the first or second half of the list. Conductance was highest at the beginning of each trial and tended to gradually decrease during that trial, as was reflected by a sig-

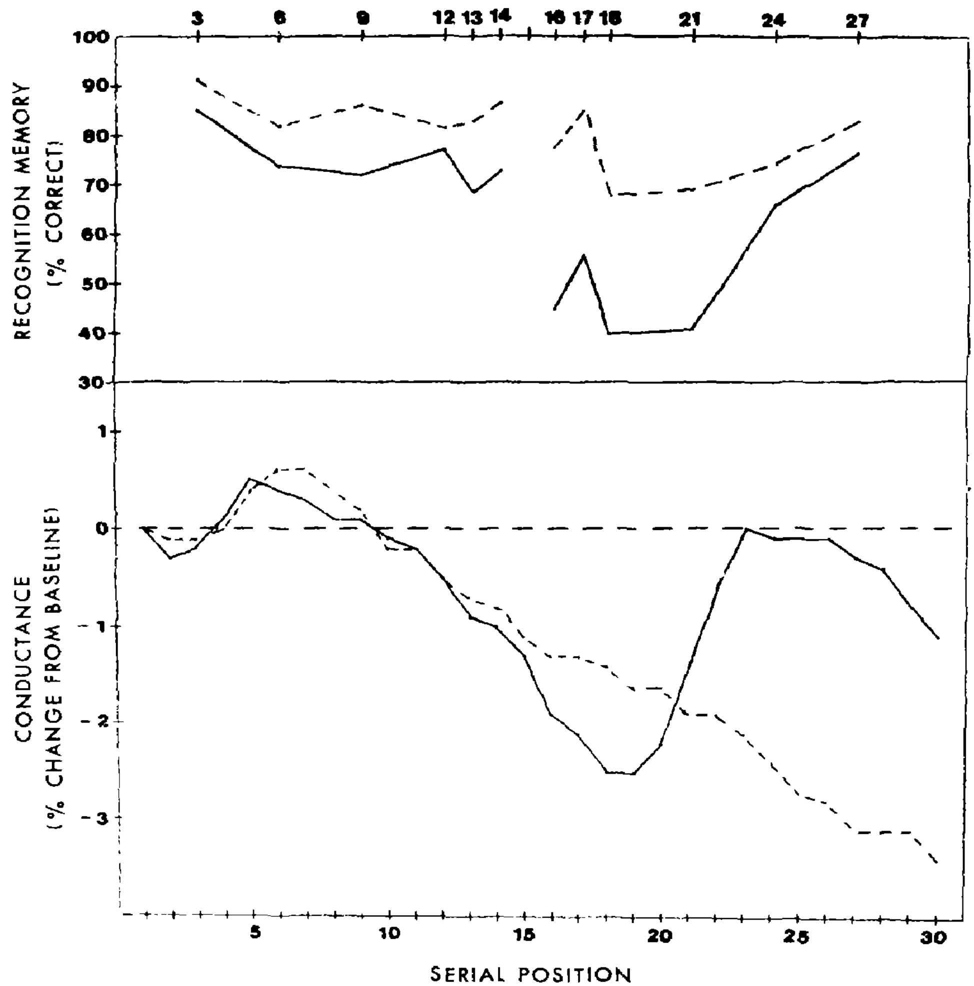


Figure 1. Percent correct recognition memory and percent change in conductance by serial position for experimental subjects (solid lines) and control subjects (broken lines)

nificant effect of serial position for the first half of the list [$F(14, 532) = 9.7, p < .01$], but not for the second half, where the response of the experimental subjects to the nude picture was reflected in an increased conductance that tended to cancel the generally decreasing conductance of the control group [$F(14, 532) = 1.6$]. The response to the critical stimulus is clearly seen in the interaction of groups by serial position [$F(14, 532) = 5.1, p < .01$]; no interaction was found for the first half of the list.

Recognition memory and palmar conductance

It is apparent, on the basis of the preceding analysis, that the presence of a nude picture had a measurable effect on the experimental subjects as measured by palmar conductance and by recognition memory. The nature and magnitude of the deficit in recognition memory replicate those found by Ellis et al. (1969) in all respects, while the autonomic measure confirms the general expectation that an 'arousing' stimulus will increase emotional arousal. It is by no means certain, however, that subjects exhibiting the largest autonomic response will also show the largest deficit in recognition memory. In order to evaluate this possible relationship between the two measures, a Pearson's r was calculated for the experimental group. For this analysis, a deficit in recognition memory was defined as the average number of items recognized at serial positions 16, 17, 18, and 21 on the seven trials containing nude items. The data on palmar conductance was taken as the difference between the maximum response to the nude item and the average value for the 3-sec period prior to the onset of that response. This revealed a nonsignificant relationship between the two measures [$r = .32$] and supports the general observation that some subjects who responded vigorously on one measure were not responsive on the other.

It was further observed that between-subject differences for latency of the autonomic response existed and that the maximum deficit in recognition memory occurred at different serial positions for different subjects. Accordingly, the autonomic responses were divided into those whose onset began within 2 sec, between 2 and 4 sec, or more than 4 sec after the critical item. The maximum deficits in recognition memory were classified as occurring at serial position 16, at positions 17 and 18, or at positions 21, 24, and 27. This analysis was concerned not with comparing the magnitude of either response but rather with determining whether or not the responses were time locked. Values of 1, 2, and 3 were assigned to the dependent measures representing early, middle, and late responses respectively. The correlation between the measures revealed no tendency for the dependent measures to bear any temporal relationship to each other [$r = .12$].

DISCUSSION

The present experiment is entirely consistent with previous findings on recognition memory (Tulving, 1969; Ellis et al., 1971; Erdelyi and Appelbaum, 1973), and it also provides evidence that a measurable autonomic (emotional) response is made to the critical item in a serial list. The

empirical generalization that a deficit in recognition memory, a behavioral response, accompanies the presence of such an item, coupled with the expectation that the critical item also elicits an emotional response, has led to an interpretation of one response in terms of the other. Similar findings have been accounted for by assuming interference with a consolidation process (Tulving, 1969). This would require the disruption of neural traces for items already encoded (retrograde amnesia) and the additional disruption of incoming signals (anterograde amnesia) by the critical stimulus. The magnitude of this effect would depend on the degree to which the critical stimulus could elicit an emotional response. The present results agree with this prediction only in a general way. On the basis of group comparisons, a large memory deficit and autonomic response were found; but when performance on the two measures was compared for individual subjects, no significant correlation was found for either the magnitude or the onset of the two measures. These findings do not support explanations of memory deficit in terms of the emotion-producing characteristics of a critical stimulus but instead suggest that while memory deficit and autonomic arousal both result from the presence of some critical item in a serial list, these two responses are largely independent. The assumption that an autonomic response influences the behavioral response was not supported. It would appear that both responses are expressions of some antecedent manipulation and are not affected by each other.

Notes

Address requests for offprints to Dennis Runcie, Department of Psychology, College of Arts and Sciences, University of Alabama, University, Alabama 35486.
Received for publication February 24, 1976; revision, May 18, 1976.

References

Detterman, D. K., and Ellis, N. R. 1972. Determinants of induced amnesia in short-term memory. *Journal of Experimental Psychology* 95:308–316.

Ellis, N. R., Detterman, D. K., Runcie, D., McCarver, R. B., and Craig, E. M. 1971. Amnesic effects in short-term memory. *Journal of Experimental Psychology* 89:357–361.

Erdelyi, M. H., and Appelbaum, A. G. 1973. Cognitive masking: The disruptive effect of an emotional stimulus upon the perception of contiguous neutral items. *Bulletin of the Psychonomic Society* 1:59–61.

Saufley, W. H., Jr., and Winograd, E. 1970. Retrograde amnesia and priority instructions in free recall. *Journal of Experimental Psychology* 85:150–152.

Tulving, E. 1969. Retrograde amnesia in free recall. *Science* 164:88–90.

Wallace, W. P. 1965. Review of the historical, empirical, and theoretical status of the von Restorff phenomenon. *Psychological Bulletin* 63:410–424.

REPORT

The Seventy-second Annual Meeting of the Society of Experimental Psychologists

The Seventy-second Annual Meeting of the Society of Experimental Psychologists was held at the University of Texas in Austin on April 9 and 10, 1976. Abram Amsel, Chairman of the Society, presided at the business meeting and at the sessions in which scientific papers were presented.

The following Members and Fellows were present: Amsel, Bevan, Bourne, Carterette, Cofer, Eriksen, Garner, Grant, Grice, Hake, D. Hurvich, L. Hurvich, Jenkins, H. Kendler, T. Kendler, Kimble, Kintsch, Liberman, Melton, Murdock, Posner, Solomon, Teitelbaum, Thompson, Tulving, Underwood.

Byron A. Campbell, I. Gormezano, Herbert M. Jenkins, David H. Krantz, and Hershel W. Liebowitz were elected to membership.

John F. Dashiel and Clifford T. Morgan died during the year. The present membership of the society is 76 Members and 54 Fellows.

The 1976 Warren Medal was awarded to Wendell R. Garner with the citation, "For his careful experimental and theoretical analysis of internal and external factors in perceptual organization."

The society accepted an invitation from Yale University and Haskins Laboratories to meet in New Haven in 1977. Wendell R. Garner and Alvin M. Liberman were elected cochairmen for the year. The executive committee consists of the Cochairmen, Gregory A. Kimble, Secretary, Philip Teitelbaum, and Abram Amsel.

G. Robert Grice, Secretary; *University of New Mexico*

AMERICAN JOURNAL OF PSYCHOLOGY
March 1977, Vol. 90, No. 1, pp. 63–72

Form class as a perceptual set

Larry G. Richards
University of Virginia

In 1972, Roydes and Osgood studied the tachistoscopic recognition of words that can be used as either nouns or verbs and found an interaction of predominant form class by set to perceive a given form class. Three experiments were done to test the generality of that finding. At luminances appropriate for reading (7 and 32 mlam), no such interaction was found, either with Roydes and Osgood's words (regardless of whether the subjects' instructions were to guess or to report accurately) or with two new lists of words that can belong to either form class.

Roydes and Osgood (1972) recently investigated the influence of set to perceive a given form class on recognition thresholds for grammatically ambiguous English words. They assembled a list of high-frequency words that could be used as either nouns or verbs. Each word, however, was predominantly used as either a noun or a verb according to the tabulations of West (1964). Subjects were pretrained to have a set to perceive either nouns or verbs or were not specially pretrained. When recognition thresholds were obtained tachistoscopically for the test words, the main effects of form class and set were not significant; however, a significant interaction of form class by set was found. In the groups pretrained to have a set to perceive nouns or verbs, thresholds were lower for words compatible with that set than for words incompatible with it.

Previous work on perceptual set for verbal material has been concerned with semantic distinctions or categories. If form class can function as a potent set-inducing variable, then the notion of set must be extended to include syntactic and other features. Further, the finding of Roydes and Osgood (1972) seems to contradict Wickens' (1970) work on release from proactive inhibition, work which failed to show form class as a significant encoding variable for words.

There are three troublesome aspects of the study by Roydes and Osgood. First, subjects were instructed to guess at what was being shown rather than to strive for accuracy; second, the starting duration used in

obtaining thresholds is not reported; and third, luminances are not reported. Roydes and Osgood report illuminance (.5 ftc) rather than luminance and fail to report the reflectance of the cards used in their study. Given the mean duration thresholds (186 to 317 msec) they obtained for short high-frequency words, it is probable that luminance was quite low and/or contrast of the stimulus words was poor.[1] If so, the situation involved minimal stimulus information, and thus guessing strategies would be very important. If subjects were guessing in the absence of information or with very limited stimulus information, then the interaction of form class by set that was found is hardly surprising. However, set as an aid to guessing is a theoretically distinct process from set as a facilitator of perception.

Three experiments were done to assess the generality and strength of Roydes and Osgood's finding. In the first, the pretraining to induce a set was extended, the instructions emphasized accuracy rather than guessing, and luminances in the normal range for reading were used. In the second, Roydes and Osgood's procedure was closely followed, except for luminances. In the third, new lists of words were employed, again following Roydes and Osgood's procedure.

EXPERIMENT I

METHOD

Subjects

The subjects were 30 undergraduates from introductory psychology courses at the University of Virginia. All satisfied a course requirement by participating in the study, and all had normal or corrected-to-normal visual acuity.

Apparatus

A two-channel tachistoscope (Scientific Prototype 800-F) was used to obtain the thresholds, which were measured in a dark room. The blank field of the tachistoscope was lighted, and the subject was instructed to look at it between stimulus presentations. Fixation lines were provided to guide the subject's focus to that portion of the visual field in which the words would appear. The subject displayed his own stimuli, when ready, by using a hand trigger. The blank field went off when the stimulus field was triggered. Luminance in both fields was 32 mlam as measured with a Macbeth illuminometer. Duration was varied in 5-msec steps starting at 5 msec. The ascending method of limits was used for each word. The criterion was two successive correct recognitions, and the duration of the first of these was taken as the recognition threshold.

Stimuli

The 40 words used by Roydes and Osgood were also used in this study. Ten unambiguous nouns were used to induce a set to perceive nouns (*noun set*), and

ten unambiguous verbs were used to induce a set to perceive verbs (*verb set*). Thresholds were obtained for 20 ambiguous words that could be used as nouns or verbs, 10 of them used predominantly as nouns (*high nouns*) and 10 of them as verbs (*high verbs*). All words were typed in lowercase pica characters.

Procedure

Each subject was randomly assigned to one of three conditions upon arrival at the laboratory. These conditions were a group pretrained to have a noun set, a group pretrained to have a verb set, and a control group. In order to maximize the influence of set, the following steps were taken in the groups pretrained to have a noun or verb set. First, subjects were informed that throughout this experiment they would be dealing with common English nouns (or verbs). Second, subjects were given a list of 10 *un*ambiguous nouns (or verbs) and asked to pronounce each word and note its form class. Third, each of the subjects then proceeded through the list a second time repeating each word and using it in a sentence. Fourth, the subjects were familiarized with the tachistoscope, and thresholds were obtained for two practice words, both of which were nouns (or verbs). Fifth, the subjects were then told that the test phase of the experiment was underway and that thresholds would now be obtained for words of the appropriate form class. Recognition thresholds were then taken for 26 words. The first 6 of these were unambiguous nouns or verbs from one of the above lists, after which the 20 ambiguous words were presented in a different random order for each of the subjects.

The control group received no pretraining. For the tachistoscopic phase, they were told that common English words would be presented. One of their practice words was an unambiguous noun and the other an unambiguous verb; the first 6 test words were 3 nouns and 3 verbs in random order. Then, the 20 ambiguous words were presented in a different random order for each of the subjects.

Subjects in all groups were instructed to report only what they saw: either "nothing" or "a blur," or particular letters or features, or total words. The experimenter recorded each subject's responses verbatim.

RESULTS

The basic data for each subject were 20 scores representing his recognition threshold for each of the ambiguous words. The pattern of mean recognition thresholds as a function of set and form class, displayed in Table 1, resembles that obtained by Roydes and Osgood, except that their mean recognition thresholds were about ten times as great as ours. In both studies, there was a trend for the group pretrained to have a noun set to show lower means than the control group. Within the group pretrained to have a verb set, high verbs were recognized faster on the average than high nouns; while the reverse was true for both of the other groups. However, the range between the highest and the lowest of these means is only 5.15 msec.

Table 1. Mean recognition thresholds (msec) as a function of set and forr class; Experiment I

	Noun set	Control	Verb set
High nouns	20 60	22 00	25.85
High verbs	21.40	23 15	24.60

For each subject, a total score was computed for both the high noun and the high verbs, and these scores were subjected to an analysis o variance. Set and form class were both fixed factors, and subjects, a random factor, was nested within levels of set. Thus, set was a between-subjects factor and form class was a within-subjects factor. Neither of th main effects even approached significance [for set, $F(2, 27) = 1.03$; fo form class, $F < 1$], nor did the interaction [$F(2, 27) = 1.04$]. In particular, the F for the interaction could have resulted by chance more than 2 times in 100 given that there were no effects due to the interaction o set with form class.

The means over the 10 subjects in each group for each of the hig nouns and high verbs were examined. Roydes and Osgood's initial hypothesis predicted changes for these words from the level of the contro group as a function of the set the subjects were pretrained to have. Considering any change from the control group, regardless of magnitude, on obtains the changes indicated in Table 2. For high nouns, the change were as Roydes and Osgood would have predicted: thresholds tended t decrease if the subject had a noun set and to increase if he had a ver set. However, for the high verbs, a verb set led to more words with increased thresholds, while a noun set led to either a decrease or no change Given a noun set, thresholds tended to decrease for words in both form classes; given a verb set, they tended to increase in both classes.

Three sets of supplementary data[2] suggested that the words *name, view can, bear, pay* and *set* were sufficiently deviant from the 9:1 or 1:9 ratio of usage as noun to verb of Roydes and Osgood to exclude them from the data. When these words are omitted, the recomputed means reported in Table 3 result. The pattern of means is essentially the same as that shown in Table 1. Selection of only those words at extreme ratios of usage does not increase the mean differences in recognition thresholds in fact, it decreases them.

EXPERIMENT II

Roydes and Osgood had encouraged their subjects to guess even if they

Table 2. Changes in mean recognition thresholds for the groups pretrained to have a noun or verb set in comparison to the control group; Experiment I

	High nouns		High verbs	
	Noun set	Verb set	Noun set	Verb set
Increase	3	8	1	7
Decrease	7	1	5	3
No change	0	1	4	0

Table 3. Revised mean recognition thresholds (msec) as a function of set and form class; Experiment I

	Noun set	Control	Verb set
High nouns	20 88	22.06	25 88
High verbs	21.08	22 33	23.50

were uncertain of the word presented; the instructions in Experiment I just reported emphasized to the subjects accurate reporting of what was seen. This difference could account for the difference in results. Accordingly, Experiment II used Roydes and Osgood's instructions to obtain guessing thresholds. Their words and procedure were also used.

METHOD

Subjects

The subjects were 30 undergraduates from introductory psychology classes at the University of Virginia. All had normal or corrected-to-normal visual acuity and were randomly assigned to experimental conditions on arrival at the laboratory.

Stimuli and procedure

The same words, procedure (including that for pretraining), and instructions used by Roydes and Osgood were used here. All words were typed in uppercase letters on index cards In each condition, four practice words were used to absorb the major practice effects usually found in studies of tachistoscopic recognition. In the two conditions in which subjects were pretrained to have a set, the practice words were of unambiguous form class corresponding to the set; in the control condition, two nouns and two verbs were used as the practice words. Durations of stimulus presentation were varied in 5-msec steps starting at 5 msec and ending at 100 msec if a word was unrecognized at that duration. Subjects were instructed to try to *guess* the identity of the word being presented. Luminance of each field was 32 mlam.

RESULTS

The means for each experimental condition and predominant form class are shown in Table 4. In all conditions, high verbs were recognized slightly faster than high nouns. Thus, the interaction reported by Roydes and Osgood did not occur under these conditions. For each subject, a total score for the 10 high nouns and one for the 10 high verbs was computed and used as basic data for an analysis of variance. The design involved one within-subjects factor (form class) and one between-subjects factor (set). The main effect of form class was marginally significant [$F(1, 27) = 7.36, \alpha \leq .05$], but neither the main effect of set nor the interaction of set by form class was significant.

EXPERIMENT III

Experiment III was done using two different lists of ambiguous words. In addition, lower luminances were employed in both fields of the tachistoscope. Roydes and Osgood's procedure and instructions were also used here.

METHOD

Subjects

The subjects were 30 University of Virginia undergraduates in psychology courses. All had normal or corrected-to-normal visual acuity.

Stimuli

Two lists of words, ambiguous as to form class, were selected from West's (1964) lists. Words showing at least a 7:3 or 3:7 ratio of usage as noun to verb and an A or AA frequency of occurrence according to Thorndike and Lorge's (1944) word count were selected. It was necessary to relax Roydes and Osgood's ratio for selection because not enough high-frequency words with a 9:1 or 1:9 separation could be found. It also proved necessary to include a wider range of lengths.[3] Ten test words were randomly selected from each dominant form class for each list.

Table 4. Mean recognition thresholds (msec) as a function of set and form class; Experiment II

	Noun set	Control	Verb set
High nouns	23 80	27.05	28 35
High verbs	22 90	26 10	25.45

Procedure

The procedure and instructions of Experiment II were also followed here. Five subjects were used in each experimental condition for each list. Luminance was 7 mlam for the blank field and 8 mlam for the stimulus field. These luminances are characterized by Boynton as the "luminance of a highly reflecting diffuse surface in feeble interior lighting" (1966, p. 289).

RESULTS

The means for each of the two lists by predominant form class under each condition appear in Table 5. For neither list was an interaction of set by form class apparent.

For each subject, scores from the 10 words in each form class were summed, and these sums were the basic data for the analyses of variance. Separate analyses for the two lists revealed no significant effects.

Figure 1 is a summary of the results of the present three experiments and those of Roydes and Osgood. Each of the plotted means is an average, taken over 10 subjects, of the individual form-class means computed over 10 words. The standard errors are computed from the form-class means of each subject. The sample size for each of the present three experiments was exactly the same as that of Roydes and Osgood ($N =$ 30, 10 subjects in each of three conditions).

DISCUSSION

The results from these three experiments limit the generality of Roydes and Osgood's finding. In the range of luminances used in the present studies, the interaction of set by form class did not occur with either type of instructions, those stressing accuracy or those stressing guesses. The range of luminances used here includes values representing normal to

Table 5. Mean recognition thresholds (msec) as a function of set and form class; Experiment III

	Noun set	Control	Verb set
List 1			
High nouns	23.1	19.70	21.8
High verbs	23.1	22.50	25.1
List 2			
High nouns	17.9	22.5	18.1
High verbs	17.6	22.4	16.1

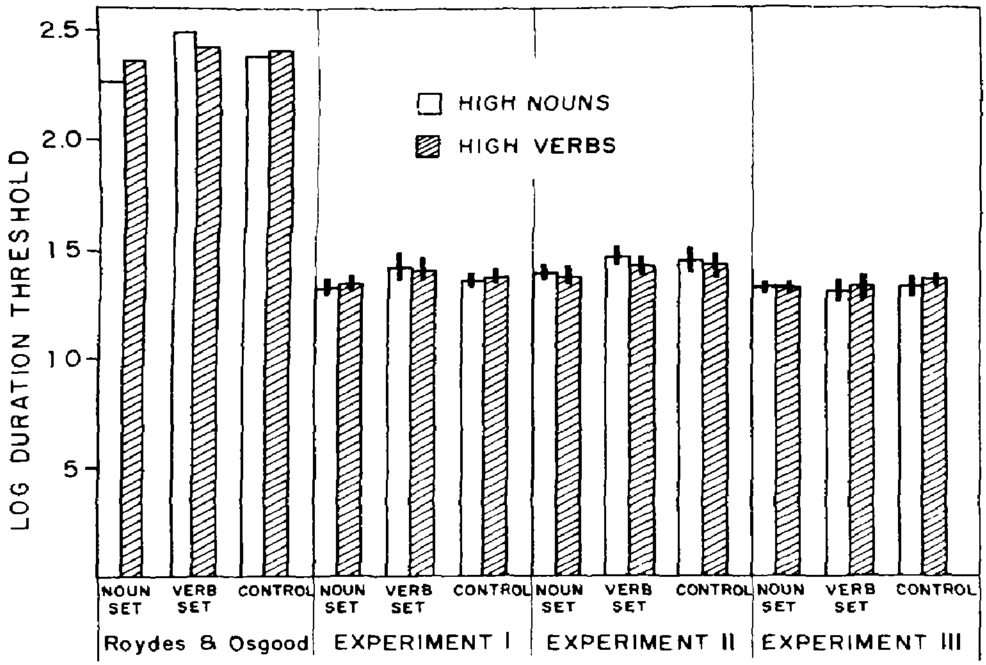


Figure 1. Log duration thresholds by group for each form class from four experiments. The solid bars indicate the standard error associated with each mean. Roydes and Osgood instructed subjects to guess at the identity of the stimulus; luminances were not reported, but stimulus contrast must have been poor. Experiment I used the same words as Roydes and Osgood; the luminance of both fields was 32 mlam, and extended pretraining to induce a set and instructions stressing accuracy were used. Experiment II was an exact replication of Roydes and Osgood's, except that the luminances were 32 mlam. Experiment III involved new words; the luminances for the stimulus and blank fields were lower (7 and 8 mlam respectively) and the instructions were to guess

poor lighting conditions for reading. Such values have been used in most studies of word recognition (at least, those which report luminance). The scale of duration used here should have enabled any potent effects of set to emerge.

In Experiment I, with instructions stressing accuracy, no effects due to predominant form class or to set were found. In Experiment II, with instructions encouraging the subjects to guess, the effect of form class was significant; note, however, that words were nested within form class, so that this effect may be due to the particular words used rather than to form class per se. The results of Experiment III, in which no effects of form class were found with different lists of words, support this conten-

tion. None of the three experiments found the interaction Roydes and Osgood reported. The presence of such an interaction in Roydes and Osgood's study might involve the resemblance of their situation to one of pure guessing, or Type I error.

Roydes and Osgood (1972) and Osgood and Hoosain (1974) interpreted the interaction of form class by set in perceptual terms, but the present results suggest that the effect was more likely due to response bias in guessing. Hochberg (1956) has conceptualized a continuum of responses ranging from complete dependence on stimulus information to complete independence of it. Roydes and Osgood's result would seem to show the effect of set as a limiter of guesses in a situation with minimal stimulus information. If the subject has little or no stimulus information but has been told that the words being shown are verbs (or nouns), then he will guess words that are predominantly verbs (or nouns). If he can determine that short words (three and four letters) are being presented, then his guesses are more limited. If he can determine the first letter of the stimulus word, that limits his guesses still further. But to say that a subject will guess verbs given a verb set and minimal stimulus information in no way implies that the set influences the perceptibility or recognizability of the stimuli. Semantic or syntactic information does not seem to influence perceptual processing in this type of task. When the perceptual information is sufficient, word recognition occurs regardless of the set induced or the type of instructions given. This hardly supports Osgood and Hoosain's contention (1974, p. 182) that Roydes and Osgood's 1972 experiment "does seem to provide unambiguous evidence for central perceptual bias."

Notes

The author thanks Philip D. Burkholder and Rick Heller for their assistance in collecting data for this paper, and L. Starling Reid and Eugene Lovelace for their comments on the manuscript. Requests for offprints should be addressed to Larry G. Richards, Department of Psychology, 102 Gilmer Hall, University of Virginia, Charlottesville, Virginia 22901. Received for publication March 25, 1976.

1. Using Turkish words with pretraining, Baker and Feldman (1956) found luminance thresholds in the range from .1 to 1 mlam when duration was fixed at 200 msec.

2. Summaries of these studies are available on request.

3. The necessity to relax Roydes and Osgood's ratio for selection to perform this study with a new set of words itself further limits the generality of their finding.

References

Baker, K. E., and Feldman, H. 1956. Threshold luminance for recognition in relation to frequency of prior exposure. *American Journal of Psychology* 69:278–280.

Boynton, R. M. 1966. Vision. In *Experimental methods and instrumentation in psychology*, ed. Joseph B. Sidowski. New York: McGraw-Hill.

Goldiamond, I., and Hawkins, W. I. 1958. Vexierversuch: The log relationship between word-frequency and recognition obtained in the absence of stimulus words. *Journal of Experimental Psychology* 56:457–463.

Hochberg, J. 1956. Perception: Toward the recovery of a definition. *Psychological Review* 63:400–405.

Osgood, C. E., and Hoosain, R. 1974. Salience of the word as a unit in the perception of language. *Perception and Psychophysics* 15:168–192.

Roydes, R., and Osgood, C. E. 1972. Effects of grammatical form-class set upon perception of grammatically ambiguous English words. *Journal of Psycholinguistic Research* 1:165–174.

Thorndike, E. L., and Lorge, I. 1944. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University.

West, M. 1964. *A general service list of English words.* London: Longmans, Green.

Wickens, D. D. 1970. Encoding categories of words: An empirical approach to meaning. *Psychological Review* 77:1–15.

AMERICAN JOURNAL OF PSYCHOLOGY
March 1977, Vol. 90, No. 1, pp. 73-83

Auditory perception of spatiotemporal patterns

Frank J. Tolkmitt and Robin Brindley
University of Melbourne, Australia

To test the tendency of subjects to perceptually organize discrete temporal patterns with regard to runs of identical stimulus events, spatiotemporal patterns of white noise were presented for reproduction. The subjects in Experiment I began their reproductions with the true starting points of the patterns, but the subjects in Experiment II, who could not perceive the true starting points, began their reproductions with either long or short runs. It is suggested that changes in runs of auditory patterns are perceptually analogous to changes in contours of visual patterns.

Investigation of the auditory perception of patterns has been negligible compared with that of the visual perception of patterns. This might be due to the difficulties that one encounters when trying to define what exactly constitutes an auditory pattern. Visual patterns can be defined in terms of features like lines, angles, contours, and so on that stimulate the retina simultaneously. Irrespective of what the constituent features of auditory patterns are, they stimulate the cochlea successively. In general, then, visual patterns are *spatial* in nature, whereas auditory patterns are *temporal* in nature.

Frequently, investigators of auditory perception use patterns that are sequences of discrete events, where each event is chosen from a number of predetermined stimuli (Royer and Garner, 1966; Handel and Lewis, 1970). If adjacent events within a pattern consist of the same stimulus, such a sequence is called a run (Royer and Garner, 1970). Normally, temporal patterns with a given number of events are played repeatedly and without a break to subjects, whose task it is to reproduce them. Under continuous repetition, the number of possible reproductions can be as great as the number of events in the patterns. Preferential reproduction is assumed to reflect the organizational tendencies that govern the perception of the subjects. The main finding is that subjects tend to perceive the runs in any given pattern as units and to place these units either at

the start or the end of their reproduction (Preusser, Garner, and Gottwald, 1970).

It could be argued that this perceptual organization is a function of the stimulus material and does not necessarily reflect a fundamental auditory principle through which all incoming acoustic information is filtered. In most studies on perception of temporal patterns, the stimuli have differed either in tone (Royer and Garner, 1970) or duration (Nazarro and Nazarro, 1970), either of which makes it easy for a subject to rhythmically group their successive occurrence. It was therefore decided to test whether or not features like runs would emerge if one stimulus (white noise) was used and its spatial occurrence, rather than tone or duration, was varied. The spatial layout for the stimulus' occurrence was identical to that used by Tolkmitt (1970), who had placed eight loudspeakers in a circle, each speaker the same distance from the head of the subject. The eight speakers could be activated in any desirable order.

It was expected that during successive and continuous activation of all eight speakers, sequences of adjacent speakers would be perceived as runs that the subjects would place either at the start or the end of their reproductions.

EXPERIMENT I

In all but two (the C patterns) of the *spatiotemporal* patterns that were used, the successive activation of adjacent speakers with white noise was interrupted at two points. At the end of a long run, the white noise either jumped one speaker, returned to it, and jumped into the long run (the A patterns) or jumped two speakers, covered them in a backward direction, and then jumped into the start of the long run (the B patterns). The jumps were never across the aural plane but always across the median plane. This was done because of the binaural cues that mediate the localization of sound. They carry information about the lateral position of the source of a sound but fail to inform us whether it is in the front or the back. Because of this ambiguity, jumps across the aural plane could have been missed and were therefore avoided.

To guard against a biased perception of any one of the patterns, they were all started at each of the eight speakers in repeated presentations. However, Royer and Garner (1970) reported that the starting point influences the perceptual organization of temporal patterns as indicated by the reproductions. Only when reproductions are independent of the starting point can they reflect perceptual organization. If they are a function

of the position of the initial speaker, it could mean that the subjects try to copy the patterns as accurately as possible, even down to the starting point, or it could indicate that it is easier for the subjects to organize the patterns around the starting point than to rely on internal organizational principles.

In an attempt to clarify this issue, two different instructions were used. One made no reference to the starting point; the other informed the subjects that an accurate reproduction did not have to start at the initial speaker. In the absence of an effect of starting point, it was expected that the subjects would begin their reproductions with a long run. This would support the notion that irrespective of stimulus material, long runs tend to guide the perceptual organization of auditory patterns.

METHOD

Subjects

Ten female undergraduates served as paid subjects. In no case did the difference in the aural sensitivity of a given subject's ears exceed 10 db (SL), over a frequency range from 250 to 8,000 Hz.

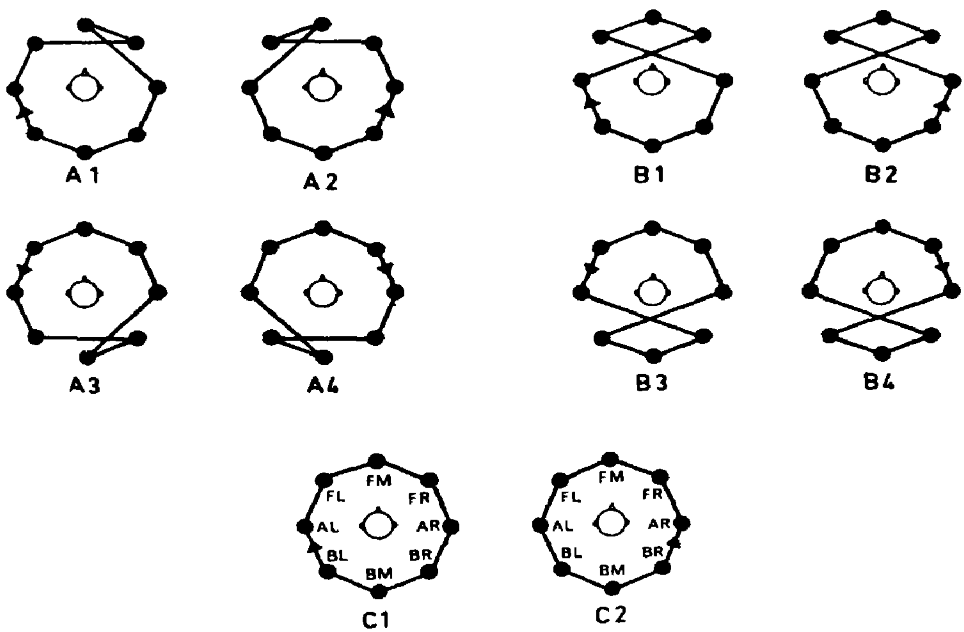


Figure 1. Structure of the spatiotemporal patterns

Apparatus

The experiment was performed in a sound-reduced chamber. Eight 20-cm loudspeakers (Kaltro FR8) were installed equidistantly at 45-deg intervals on a circular supporting frame of 3 m in diameter. The speakers were matched with regard to their output characteristics over a frequency range from 50 to 20,000 Hz. Each speaker was driven by its own amplifier and had its own volume control. The output from the amplifier was gated into the speakers by means of eight magnetic relays (Dual-Incline Pack, RA 3038) with a .15-msec switching time. The relays were interfaced to a PDP-11 computer. The eight amplifiers were driven by a white-noise generator built in the departmental workshop, and the eight speakers were calibrated to produce white noise of 50 db (SPL) at the position of the subject's head.

For purposes of reference, the eight speakers (their positions) are labeled front median (FM), front right (FR), aural right (AR), back right (BR), back median (BM), back left (BL), aural left (AL), and front left (FL); see patterns C1 and C2 in Figure 1.

Design

The three variables under investigation were instructions, pattern, and position. The ten subjects were divided into two groups of five each, and the two groups received different instructions. The instructions to both groups called for accurate reproduction of the spatiotemporal patterns; however, while the subjects in the group with *specific* instructions were told that the term 'accurate' did not imply that the starting point of the presented pattern and its reproduction had to be the same, this instruction was not given to the subjects with *nonspecific* instructions.

Pattern and position were treated as within-subject variables. The ten patterns that were used are illustrated in Figure 1. The dots represent the spatial location of each of the eight speakers (position) around the subject's head and the connecting lines indicate their temporal sequence, in the direction shown by the arrows. Each pattern was without an inherent beginning or end and, when repeated continuously, could be started at any one of the eight positions without affecting its spatiotemporal structure. Eight of the patterns had two jumps and a long and a short run. The jumps always occurred in a lateral direction in the front or back of the subject in order to aid their detection. These eight patterns were derivatives of two basic spatiotemporal sequences (A and B) reflected along the median and aural planes. The remaining two patterns (C1 and C2) were unstructured and served as controls for preferential starting points.

All patterns were presented eight times in a random order, and each one was started once from each of the eight positions. The combination of these two variables yielded 80 trials (10 patterns × 8 positions), which were randomly presented under the control of a PDP-11 computer. Each trial was presented repeatedly at a rate of 800 msec per speaker, and the time for a full pattern was therefore 6.4 sec. This rate had been determined as optimal for pattern identification in a pilot study.

The subject sat in the middle of the circle of speakers, her head in the intersection of the median and aural planes, and she had to face the front-median speaker. She held on her lap a board with a start, stop, and erase button, as well as eight buttons arranged in a circle and corresponding to the positions of the

eight speakers. By means of the buttons the subject started the trials and stopped them when she felt confident that she could reproduce their sequence on the eight buttons for position. If she made a mistake, she could erase the reproduction and respond again. In case a trial was incorrectly reproduced, it was presented again at a later stage. After 80 correct reproductions, the computer terminated the experiment and printed out the raw data. They consisted of two response indices for each trial: its presentation time (*latency*) and the starting point of the subject's reproduction (*subjective starting point*).

RESULTS

The two types of responses, latency and subjective starting point, are thought to indicate different behavioral aspects. The latencies presumably reflect the perceptual difficulties the subjects experience with the various patterns, while the subjective starting points presumably reflect the perceived structure of the patterns.

Latencies

Latencies ranged from 8.00 to 50.00 sec and were clearly a function of pattern (A1, 14.1; A2, 14.5; A3, 15.9; A4, 15.6; B1, 22.2; B2, 21.7; B3, 22.2; B4, 20.8; C1, 11.1; and C2, 11.4 sec). Neither one of the remaining two factors seemed to have affected the latencies. This was confirmed by an analysis of variance for repeated measures, which produced a main effect for pattern only [$F(9, 12) = 25.5$]. Apparently, the C patterns were easiest to perceive, followed by the A patterns and then by the B patterns. In terms of their structures, this can be interpreted to mean that decreasing the longest run increases the perceptual difficulties.

Subjective starting points

For each type of instructions, a contingency table was made of the absolute frequencies with which each position was chosen as the starting point for the reproduction of the ten patterns. The averaged frequencies of these subjective starting points for the three groups of patterns (A, B, and C) are listed separately for the two instructional groups in Table 1. As there were eight speaker positions and each pattern was started once from each one of them, the relative frequencies of the subjective starting points should all be close to .125 if they were a function of position only. Systematic deviations from this value would indicate that the subjects structured the patterns. If the subjects tended to start the experimental patterns (A and B) at either the long or short run, the relative frequencies for positions AL and AR (start of long runs) or for positions FL, FR, BL, and BR (start of short runs) should be selectively greater than .125. However, the obtained frequencies (Table 1) seem to deviate only marginally

Table 1. Relative frequencies with which positions were chosen as the subjective starting points of the ten patterns

Pat-terns	Position								% true starting points
	FM	FR	AR	BR	BM	BL	AL	FL	
Experiment I, nonspecific instructions									
A	.106	131	113	163	.088	113	163	.125	69
B	.138	100	131	131	.075	131	162	.131	77
C	.188	.075	.100	063	134	088	238	.113	.60
Experiment I, specific instructions									
A	.144	.156	138	.150	.088	.088	131	.106	54
B	.075	.166	.166	.156	.056	.094	131	.156	.54
C	.320	.088	063	.063	173	113	138	.038	.34
Experiment II									
A	.066	144	256	.116	025	.084	228	081	10
B	.009	.116	.338	044	.003	038	.347	109	13
C	.269	013	.094	050	.244	.119	.181	.031	11

from this value. Such a result tends to indicate a strong effect of position — to show that the subjects started the reproductions at the true starting points. In the last column of Table 1 the proportions of subjective starting points that fell at the true starting points for the three groups of patterns are listed separately for each experimental group. These values favor a strong effect of position, since the subjects with nonspecific instructions reproduced about two-thirds of the patterns from the true starting point while the subjects with specific instructions did the same for nearly half of the patterns.

In order to evaluate the quantitative effects of pattern and position on the subjective starting points, the data were submitted to a multivariate analysis of transmitted information (McGill, 1954). Such an analysis is appropriate in the present case, as it assesses the association between independent and dependent variables on the basis of discretely categorized data.

A separate analysis was run for each instructional group, and Table 2 gives a breakdown of the response uncertainty in terms of its association with the input variables. Pattern accounted for only a small amount of response uncertainty in both analyses, while position accounted for nearly 50% of response uncertainty with nonspecific instructions and for half that amount with specific instructions. Apparently, the two types of instructions had differential effects on the subjects' tendency to use the position of the initial speaker as a starting point for the reproduction. How-

Table 2. Analysis of transmitted information

	Nonspecific instructions		Specific instructions	
	Bits	%	Bits	%
Pattern (subjective starting point)	.14	5	.28	9
Position (subjective starting point)	1.45	49	.70	24
Interaction	.49	16	.64	22
Error	89	30	1 35	45
Total response uncertainty	2.97		2 97	

ever, it had been hoped that the specific instructions would minimize the effect of position even further, in order to make the experimental design sensitive to a possible effect of pattern.

Unless the effect of position is eliminated, the possibility of a strong effect of pattern cannot be ruled out. The two effects are mutually exclusive and only in the absence of the former can the latter manifest itself. It was therefore decided to present the patterns in such a way that the effect of position would be minimal.

EXPERIMENT II

In order to minimize the effect of position, a slightly different procedure was used. Instead of presenting the patterns at a constant rate, each pattern was introduced at a very fast rate, which was then decelerated to the rate used in Experiment I. The initial rapid succession of speakers was expected to make it impossible for the subjects to perceive the true starting point of the patterns. Apart from this procedural change, the same within-subject variables as in Experiment I were investigated, namely, pattern and position.

METHOD

Subjects

Ten female undergraduate students served as paid subjects. Their differences in aural sensitivity at the two ears again did not exceed 10 db (SL) over a frequency range from 250 to 8,000 Hz.

Design

The two investigated variables were the ten patterns and eight positions of Experiment I. The first cycle of each pattern was presented at a decelerating rate. The exact presentation durations for the first eight speakers were always 20, 40, 80, 160, 320, 480, and 800 msec. The following cycles were presented

at a constant rate of 800 msec till the subject terminated the pattern. This procedure made it impossible for the subject to locate the beginning of the pattern. Otherwise, the apparatus and procedure were exactly as in Experiment I. The subjects were instructed to reproduce the patterns on the response board. Only the subjective starting point was recorded.

RESULTS

This time the relative frequencies of the subjective starting points deviated considerably from .125 for each of the eight speaker positions (see Table 1). In the subjects' reproductions, the control patterns were started predominantly from the median-plane positions while the experimental patterns were frequently started at the aural-plane positions, which happened to be the beginnings of the long runs.

A multivariate analysis of transmitted information indicated that 46% of the response uncertainty was related to the independent variables as follows: only 1% was accounted for by position, 38% by pattern, and 7% by their interaction. The drastic change in the contribution of position and pattern lends credibility to the conjecture that perceptual tendencies take effect only when the subject is unaware of the true starting point of the pattern.

To further analyze the nature of the relationship between pattern and subjective starting point, the data for each pattern were slightly rearranged. As pointed out before, the experimental patterns were variations of two basic spatiotemporal sequences, A and B. The A patterns had a long run of six and a short of two events; the long and short runs of the B patterns consisted of five and three events. Pooling the subjective starting points at corresponding positions within the A, B, and C patterns indicated the frequencies with which each position within a respective long or short run was chosen as the starting point for the subjects' reproductions.

Figure 2 plots this data for the A, B, and C patterns of Experiment II and for their counterparts of Experiment I. While the subjects in Experiment I showed only a tendency to start the reproduction of the experimental patterns with the short run, the subjects in Experiment II showed a definite preference for starting with the long as well as the short run. Indeed, the subjects in Experiment II started their reproductions of B patterns almost exclusively with a long or short run. These data correspond well with some of the results of Royer and Garner (1966). While the probabilities of starting with a long or short run in the present Experiment II were .47 or .26 for the A sequences and .66 or .30 for the B

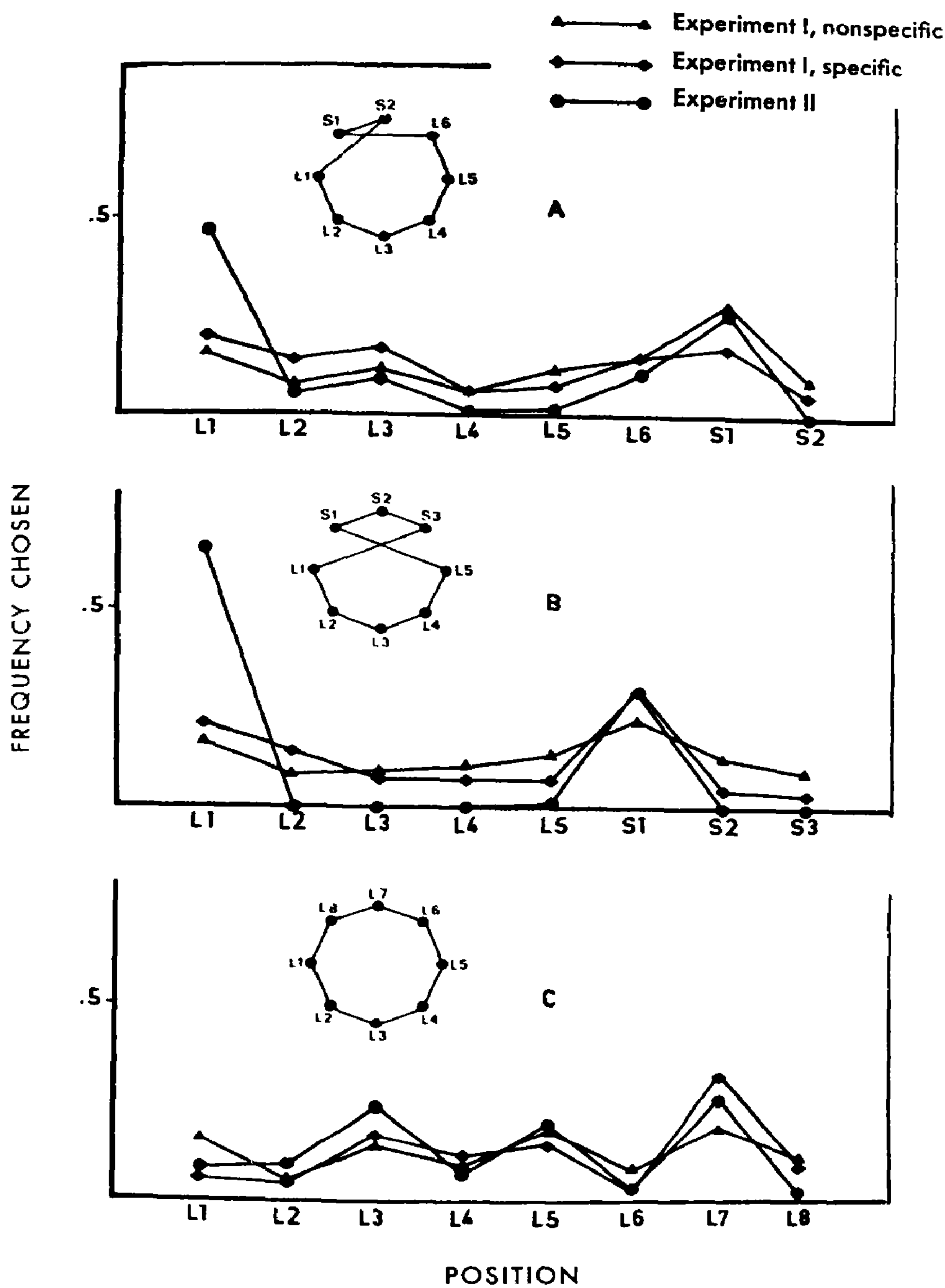


Figure 2. Relative frequencies with which corresponding positions in the long (L) and short (S) runs of the A, B, and C patterns were chosen as the subjective starting points; in Experiment I, instructions were either specific or nonspecific

sequences, Royer and Garner reported respective probabilities of .56 or .30 for structurally compatible A patterns and .54 or .38 for equivalent B patterns.

In the case of the C patterns, the subjects in Experiment II tended to start the reproductions at the median-plane positions, while of the subjects in Experiment I, only those with specific instructions seemed to have a slightly similar preference for these positions.

CONCLUSIONS

The reported perceptual organization of temporal patterns into runs (Preusser, Garner, and Gottwald, 1970) has now, in Experiment II, been observed for spatiotemporal patterns. It can therefore be concluded that this perceptual tendency is independent of stimulus material and reflects a fundamental principle in auditory perception. However, it seems to be a latent principle that manifests itself only in the absence of simpler structural aids like the true starting points of the patterns in Experiment I.

The fact that subjects anchored their perceptions at the beginnings of runs could be interpreted to mean that changes in runs convey the most structural information about discrete auditory patterns. Changes in the contours of visual patterns are known to be structurally important components, the number of independent turns in the contour of a visual pattern contributing toward its perceived complexity (Attneave and Arnoult, 1956; Vitz and Todd, 1971). Apparently, turns in contours convey more information about visual patterns than points along their continuous lines. Something much the same seems to be true for the auditory patterns used here, if one assumes that the frequencies with which positions were chosen as subjective starting points indicate the amount of structural information conveyed at those positions. Figure 2 shows that these frequencies were greatest at the positions marking changes between runs and smallest for adjacent events (positions) within runs. Thus, just as changes in spatial contours are important aspects of visual patterns, so changes in temporal contours seem to mediate the structural information conveyed by auditory patterns.

Notes

Requests for offprints should be sent to F. J. Tolkmitt, Center for Systems Neuroscience, University of Massachusetts, Amherst, Massachusetts 01002. Received for publication December 12, 1975; revision, May 17, 1976.

References

Attneave, F., and Arnoult, M. D. 1956. The quantitative study of shape and pattern perception. *Psychological Bulletin* 53:452–471.

Handel, S., and Lewis, W. E. 1970. Effects of practice on the perception of temporal patterns. *Quarterly Journal of Experimental Psychology* 22:97–108.

McGill, W. J. 1954. Multivariate information transmission. *Psychometrika* 19: 97–116.

Nazarro, J. R., and Nazarro, J. N. 1970. Auditory versus visual learning of temporal patterns. *Journal of Experimental Psychology* 84:477–478.

Preusser, D., Garner, W. R., and Gottwald, R. L. 1970. Perceptual organization of two-element patterns as a function of their component one-element patterns. *American Journal of Psychology* 83:151–170.

Royer, F. L., and Garner, W. R. 1966. Response uncertainty and perceptual difficulty of auditory temporal patterns. *Perception and Psychophysics* 1: 41–47.

Royer, F. L., and Garner, W. R. 1970. Perceptual organization of nine-element auditory temporal patterns. *Perception and Psychophysics* 7:115–120.

Tolkmitt, F. J. 1970. Auditory pattern perception: Processing limits and organizational tendencies. *Journal of Experimental Psychology* 86:171–180.

Vitz, P. C., and Todd, T. 1971. A model of the perception of simple geometric figures. *Psychological Review* 78:207–228.

'NOUNCEMENT

EB/ETS regional seminars:
tical Problems and Issues in the Education of Minorities

: College Entrance Examination Board (CEEB) and Educational Testing vice (ETS) are sponsoring eight regional seminars throughout the United tes this spring in an effort to revitalize the nation's lagging interest in the cational opportunities of minority students. A major aim of the seminars, ch have been in the planning stage since last fall, is to identify obstacles faced these students and consider actions for change. Each of the seminars on tical Problems and Issues in the Education of Minorities, to be held in a major , will focus on a critical problem or issue that traditionally has influenced the cational success of students from minority groups, notably Blacks, Mexican ericans, Puerto Ricans, and American Indians.

h CEEB and ETS note that while minority problems received serious atten- during the 1960s, the mood of the nation has shifted in the last few years concerns for minorities have been given a much lower national priority, tly in the belief that the civil rights laws of the 1960's, and the special efforts le during that period, solved the problem.

h seminar will cover one of the eight issues. And these will range from social :es that affect the education of minority students, desegregation, changing ernmental policies and attitudes at various levels, and special problems faced students with bilingual or bicultural backgrounds to use of admissions policies barrier.

: first seminar will be held at Los Angeles, April 1, 1977, and will focus on blems faced by students from bilingual/bicultural backgrounds. Other semi- s will be in Boston, April 4, on changing policies and attitudes at the federal state level; in Princeton, April 11, on social forces that affect the education ninorities; in Austin, April 15, on career counseling of minority students; in inta, April 22, on desegregation; in Cleveland, April 25, on education as a icle for rescuing urban minority youths from poverty and frustration; in iver, April 29, on educational concerns of American Indians; and New York , May 9, on admissions policies. For additional information about the inars, please contact

liam U. Harris
ıcational Testing Service
ıceton, New Jersey 08540

AMERICAN JOURNAL OF PSYCHOLOGY
March 1977, Vol. 90, No. 1, pp. 85–93

The role of number and familiarity of stimuli in the perception of brief temporal intervals

H. R. Schiffman and Douglas J. Bobko
Rutgers — The State University

The influence of stimulus number and familiarity on judged duration were investigated. The stimuli were slides with different numbers of stimulus elements of a familiar or unfamiliar nature. The task was to reproduce the durations of the slides, each shown for four intervals: 5, 9, 13, and 17 sec. The results show that the number of stimulus elements presented within a given interval affected its perceived duration, although the familiarity of those elements (as defined herein) did not. Finally, the shorter intervals sampled here were overestimated and the longer intervals were underestimated, thereby supporting Vierordt's law.

Ornstein (1969) has proposed that the perception of a temporal interval is affected by the physical character of the stimuli perceived within the interval, as those stimulus characteristics serve to determine the encoding and storage of stimulus information in memory. Ornstein's basic premise involves a computer metaphor, in that perceived duration is presumably constructed from the contents of mental storage. That is, the amount of information stored in memory and, more important, the manner in which it is encoded determine the perceived duration of a particular interval, the duration lengthening as storage size increases.

Thus, one stimulus attribute that should affect memory storage and hence influence perceived duration is the *familiarity* of the stimulus elements defining the temporal interval. By Ornstein's argument, less storage space is required for familiar than for unfamiliar stimuli, since unfamiliar objects or events (those stimuli with which an observer has little experience) do not have the advantages of ease of coding or 'chunking' that are afforded familiar stimuli. Hence, the less familiar the stimuli occurring within an interval, the longer its perceived duration. The results of several recent studies indicate that stimulus familiarity of a verbal nature affects the experience of durations as brief as 10–30 msec (Avant and Lyman, 1975; Avant, Lyman, and Antes, 1975); these experiments found

that tachistoscopic presentations of nonwords (unfamiliar stimuli) were judged longer than presentations of words or single letters. However, with the possible exception of Ornstein's indirect study (1969, Study 5, pp. 73–79), which used a well-practiced motor task (pursuit-rotor), the role of stimulus familiarity of a nonverbal nature has been the subject of little empirical research in time perception.

A second stimulus attribute proposed to influence time perception is the *number* of stimuli within the interval. By this argument, the more stimulus elements contained and perceived within a given time interval, the greater the presumed storage requirement and the longer the interval's perceived duration. Ornstein's research and that of others (e.g., Buffardi, 1971; Fraisse, 1963; Hall and Jastrow, 1886) lend support to this contention. However, in light of the fact that the primary sense modality employed in previous studies was the auditory one, the present experiment was designed to confirm and extend those findings to the visual modality.

Accordingly, the present study systematically varied the familiarity and the number of visual stimuli within an interval across a series of time intervals. It was designed to test the hypothesis, based on sparse earlier findings, that for a given temporal interval, the greater the number of elements and the less their familiarity, as defined herein, the greater the experienced duration of that interval. A point of subsidiary interest was to assess Vierordt's law (see Woodrow, 1951, p. 1225), which holds that within a series of time intervals, the shorter intervals are overestimated and the longer intervals are underestimated.

METHOD

Subjects

The subjects were 69 male volunteers from an introductory course in psychology. Although they were aware that their task was of time estimation, they were naive with regard to the experimental hypotheses.

Stimuli and apparatus

The experimental stimuli were a dozen 35-mm slides. The contents of these positive color transparencies varied in three main characteristics.

The first stimulus variable was the level of familiarity. There were three conditions of stimulus familiarity: those with *homogeneous familiar* stimuli, in which all the elements on a slide were the same (e.g., household keys); those with *heterogeneous familiar* stimuli, in which the elements on a slide were familiar but different objects (e.g., thumbtack, light bulb, paper clip); and those with *heterogeneous unfamiliar* stimuli, in which the elements on a slide were unfamiliar and different objects (e.g., electrical connector, capacitor, opti-

cal filter). Specifically, there were four slides at each level of familiarity. In terms of their relative information value and ease of chunking, the following relationships were assumed: heterogeneous unfamiliar > heterogeneous familiar > homogeneous familiar, where the condition with heterogeneous unfamiliar stimuli provided the most information and required the largest storage.

The second stimulus variable was the number of elements on a given slide. At every level of familiarity, each of the four slides contained a different number of elements, either *three, five, seven,* or *nine* elements.

The third stimulus variable was the length of the interval for which the slides were presented. Each slide was presented for intervals of *5, 9, 13,* and *17* sec.

The mode of stimulus presentation was by a projector casting a 62-by-43-cm image at eye level on a 122-by-91-cm rear projection screen located at 1.2 m from the subject. The timing mechanism of the projector was modified and wired in circuit with an interval timer so that each stimulus slide could be presented for the specified interval. The interval was defined by the duration of the exposure of the stimulus slide. When the interval was ended, the projection screen went dark.

The responses — estimated durations — were made by the method of reproduction. A microswitch connected with a digital stop clock was mounted beside the subject on a small table. Depression of the switch started the clock and release of the switch stopped the clock. Before the experimental trials, each subject was instructed in using the switch to produce his responses.

Room lighting during the experimental sessions was about 3 cd m^{-2}. All control and recording apparatus were in a room adjacent to the experimental testing room; in addition, a low setting of white noise was used to mask extraneous sound.

Procedure

Before the experimental testing, the subject was seated before the projection screen, his watch was removed, and the following instructions were read. "You are about to participate in an experiment involving memory and time judgment. A number of different slides will be shown on the screen. After each slide has appeared you will be asked to depress this switch [the experimenter pointed to the microswitch] to reproduce the length of time the slide had been on. You will do this for every slide that appears on the screen. Please do not count, tap, or use any other mnemonic devices during the experiment to judge time span. Keep your attention on the objects contained on the slides. After all the slides have been shown, you will be given a short questionnaire to test your memory of the objects contained on the slides. Do you have any questions?"

Two practice stimuli, slides showing one and two keys, were presented for 7 and 15 sec respectively to each subject before the experimental sessions began.

Design

The subjects were randomly assigned to one of three conditions of stimulus familiarity, resulting in three independent groups of 23 subjects each. Within each of these conditions, every number of elements (three, five, seven, and nine) was paired with every length of interval (5, 9, 13, and 17 sec), for a total of 16 treatments; in addition, each treatment was presented three times, for a total of 48 estimated durations per subject. In other words, each subject saw

only four slides at a given level of familiarity, but each slide was presented a dozen times (four different intervals and with three replications). The order of presentation within each condition of stimulus familiarity was random.

On completion of an experimental session, those subjects serving in the experimental conditions with heterogeneous familiar and heterogeneous unfamiliar slides were given a recognition test for the elements contained on those slides. The test consisted of a list of 30 words containing the names of those objects shown on the original slides. The dependent measure was the percent correct recognition.

RESULTS

Each estimated duration by every subject was converted to a ratio of the estimated duration of the interval to its actual duration, thus providing a direct measure of the subjects' accuracy of reproduction for each different interval presented. Ratios of 1.00 show perfect accuracy, ratios below 1.00 represent underestimation of the actual interval, and ratios above 1.00 indicate an overestimation of the interval (see Hornstein and Rotter, 1969, for a further discussion of such ratio transformations).

The relevant findings are presented in Table 1, which shows the mean ratio and standard deviation for each treatment. As indicated, there was a clear and systematic change in the variability of mean ratios as the actual interval changed; specifically, the variability of estimated duration decreased as the length of the interval increased. However, in a situation such as this with equal and large cell sizes, moderate deviations from homogeneity of variance do not seriously affect the sampling distribution of the F statistic (see Scheffé, 1959, p. 354; Cochran, 1947). Consequently, a three-way analysis of variance with repeated measures on actual interval and number of elements (Winer, 1962, pp. 539–545) was done using mean ratios of the three estimated durations made by each subject for every treatment. The results indicated a significant main effect of number of elements [$F(3, 198) = 8.63, p < .01$]. A Newman-Keuls test on the mean estimated duration for the four levels of number of elements was performed: it indicated that estimated durations for seven elements did not differ from those for five or three elements but that those for slides containing nine elements differed from those for seven, five, and three elements and that those for five elements differed from those for three [each at $p < .01$]. Thus, the hypothesis of a positive relationship between number of elements and perceived (experienced) duration was in general supported.

The main effect of actual interval was also significant [$F(3, 198) = 38.32, p < .01$]. The means for the four intervals presented all differed significantly from each other [$p < .01$].

Table 1. Mean ratios (and standard deviations) of estimated durations by number of stimulus elements and level of familiarity

	Three elements				Five elements				Seven elements				Nine elements			
	Interval (sec)				Interval (sec)				Interval (sec)				Interval (sec)			
	5	9	13	17	5	9	13	17	5	9	13	17	5	9	13	17
Homogeneous familiar stimuli																
	1.10	98	99	.88	1 17	1 04	.99	.87	1 07	1 11	.92	.86	1.20	1.16	.98	.91
	(28)	(23)	(20)	(20)	(36)	(33)	(24)	(.16)	(30)	(24)	(.16)	(.16)	(.32)	(.30)	(.16)	(.13)
Heterogeneous familiar stimuli																
	1 17	92	89	86	1 19	1 03	99	85	1.06	1 10	1.00	.80	1.21	1.21	1.00	.91
	(34)	(17)	(21)	(.18)	(36)	(.24)	(24)	(.15)	(.29)	(23)	(.21)	(.22)	(.35)	(.23)	(.20)	(.16)
Heterogeneous unfamiliar stimuli																
	1 08	99	95	97	1.17	1 03	95	.90	1 12	1 04	.96	92	1 18	1.14	.98	.94
	(.38)	(.17)	(16)	(16)	(.37)	(23)	(.17)	(.15)	(.57)	(.27)	(.19)	(.19)	(43)	(.29)	(.20)	(.16)

The main effect of stimulus familiarity [$F(2, 66) = .06$] was not significant. However, the interaction between number of elements and actual interval was significant [$F(9, 594) = 3.53, p < .01$]. The interaction effects are shown in Figure 1, where mean ratio scores, collapsed over stimulus familiarity, are plotted by interval for the four levels of number of elements: as indicated, the functions for nine and seven elements are not parallel to those for five and three elements; that is, the rate of change in estimated duration from 5 to 9 sec was less for nine and seven elements than for five and three elements. Additionally, Figure 1 shows an intersection of the function for three elements with the functions for both seven and five elements. No other interactions were significant. Finally, as Figure 1 illustrates, these findings confirm Vierordt's law: within the series of intervals employed here, the shorter intervals were overestimated and the longer intervals were underestimated.

The results of the recognition test indicated that subjects in the condition with heterogeneous familiar stimuli recognized correctly more of the elements contained on their slides than did subjects in the condition with heterogeneous unfamiliar stimuli [$t(43) = 5.96, p < .001$].

DISCUSSION

The results of the present experiment clearly indicate that perceived duration lengthens as the number of elements filling the interval increases, at least over the range of stimulus parameters employed here. This finding confirms the effect reported by others for auditory stimuli and extends to the visual modality the positive relationship between perceived duration and the number of events filling an interval. However, with regard to Ornstein's theory, the present findings are equivocal, in that a precise cause-and-effect relationship of stimulus familiarity and experienced duration is not indicated. In fact, a simple 'input register' (e.g., Ornstein, 1969, p. 38) accounts for the effect of the number of elements; that is, experienced duration lengthens directly as the amount of monitored incoming information increases.

The systematic errors in estimated duration evident in this study (the overestimation of the shorter and the underestimation of longer intervals) clearly support Vierordt's law. That the judgment of a given stimulus is made within the context of the range, frequency, magnitude, and so on of the other stimuli in the series has been amply demonstrated (see Helson, 1964, for a review). In fact, Vierordt's law is conceptually equivalent to the more general 'central tendency' (Hollingworth, 1910), and it suggests that the perception of stimulus events involves a contextual or judgmental

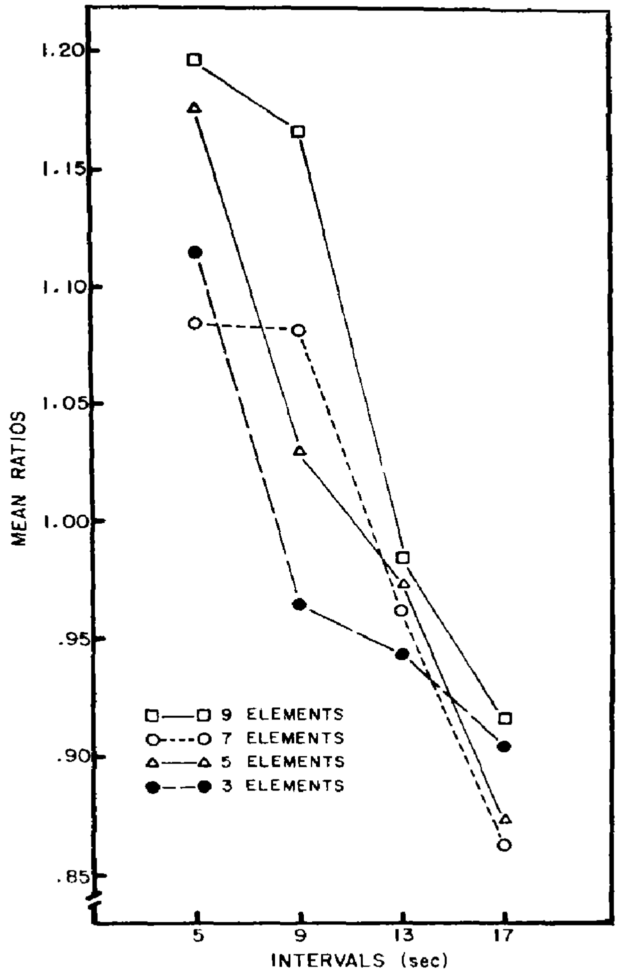


Figure 1. Mean ratios by actual intervals for number of stimulus elements (the ordinate represents the ratio of the estimated duration of the interval to its actual duration)

process. However, of interest here is the finding that the regression of estimates toward the mid-range of the stimulus series occurs in the absence of any situationally relative or restricted response mode (see, e.g., Braud and Holborn, 1966). This result is in agreement with other reports (e.g., Schiffman and Bobko, 1974; Stevens and Greenbaum, 1966), and it indicates that the central tendency operates also with a rather direct judgmental indicator (i.e., the method of reproduction).

There are several possible explanations for the failure of stimulus fa-

miliarity to affect perceived duration. Although the results of the recognition test do not support this possibility, the levels of stimulus familiarity as defined herein may not have produced sizeable differences in storage space, thereby resulting in a lack of effect of stimulus familiarity. Also, the repeated exposure of the heterogeneous unfamiliar stimuli may have reduced the amount of storage space required, the extended experience with these unfamiliar stimuli making them familiar even in the absence of a proper verbal label, perhaps through a more efficient visual code (Avant and Lyman, 1975). Finally, stimulus familiarity may have an effect only under situationally subjective or relative conditions. Specifically, perceived duration may be affected by stimulus familiarity only when each subject is exposed within an experimental session to both familiar and unfamiliar stimuli.

Notes

This research was supported by funds from the Research Council of Rutgers — The State University, Grant 07-2109, and the Busch Memorial Fund. The authors thank Philip Bobko for helpful suggestions on the statistical analyses. Received for publication January 22, 1976.

References

Avant, L. L., and Lyman, P. J. 1975. Stimulus familiarity modifies perceived duration in prerecognition visual processing. *Journal of Experimental Psychology* 104:205–213.

Avant, L. L., Lyman, P. J., and Antes, J. 1975. Effects of stimulus familiarity upon judged visual duration. *Perception and Psychophysics* 17:253–262.

Braud, W. G., and Holborn, S. W. 1966. Temporal-context effects with two judgmental languages. *Psychonomic Science* 6:151–152.

Buffardi, L. 1971. Factors affecting the filled-duration illusion in the auditory, tactual, and visual modalities. *Perception and Psychophysics* 10:292–294.

Cochran, W. G. 1947. Some consequences when assumptions for the analysis of variance are not satisfied. *Biometrics* 3:22–38.

Fraisse, P. 1963. *The psychology of time*. New York: Harper & Row.

Hall, J. S., and Jastrow, J. 1886. Studies of rhythm. *Mind* 11:55–60.

Helson, H. 1964. *Adaptation-level theory*. New York: Harper & Row.

Hollingworth, H. L. 1910. The central tendency of judgment. *Journal of Philosophical Psychology* 7:461–469.

Hornstein, A. D., and Rotter, G. S. 1969. Research methodology in temporal perception. *Journal of Experimental Psychology* 79:561–564.

Ornstein, R. E. 1969. *On the experience of time.* Middlesex, England: Penguin Books.

Scheffé, H. 1959. *The analysis of variance*. New York: Wiley.

Schiffman, H. R., and Bobko, D. J. 1974. Effects of stimulus complexity on the perception of brief temporal intervals. *Journal of Experimental Psychology* 103:156–159.

Stevens, S. S., and Greenbaum, H. B. 1966. Regression effect in psychophysical judgment. *Perception and Psychophysics* 1:439–446.

Winer, B. J. 1962. *Statistical principles in experimental design.* New York: McGraw-Hill.

Woodrow, H. 1951. Time perception. In *Handbook of experimental psychology,* ed. S. S. Stevens. New York: Wiley.

ANNOUNCEMENT

Fourth biennial congress of the
International Society for the Study of Behavioural Development

The International Society for the Study of Behavioural Development (ISSBD) will hold its fourth biennial congress at the University of Pavia, Italy, on September 19–24, 1977.

The general theme is "Biosocial Aspects of Development: An Interdisciplinary Approach." It includes 20 symposia and several lectures. The topics to be covered by the symposia are kinship genetics in primates and human development; genetic aspects of development; children at risk; nutrition and mental development; the nexus of relationships surrounding the growing child; biological and social aspects of sex differences; the influence of urban stress on biosocial development; language development, normal aspects; language development, pathological aspects; neurophysiological correlates of cognitive development; cognitive development; stages in thinking; biological and psychological aspects of sexuality; cultural transmission of value; methods and strategies for research training in human development; critical and sensitive periods in human growth and development; medical and psychological aspects of aging; the problems of retirement age; models of development; and antecedents of deviance. For further information, write

Secretary, IV Binnial Congress
International Society for the Study of Behavioural Development
Institute of Psychology of the Faculty of Letters and Philosophy
University of Pavia
Corso Strada Nuova, 65
27100 Pavia, Italy

AMERICAN JOURNAL OF PSYCHOLOGY
March 1977, Vol. 90, No. 1, pp. 95-102

Depriving rats of REM sleep: The identification of a methodological problem

Robert A. Hicks, Arlene Okuda, and Dianne Thomsen
San Jose State University

Several studies that used Jouvet's platform-in-the-water technique to deprive rats of REM sleep are reviewed. It is observed that a consistent feature of the design of these studies was to ignore the ratio between the size of the animal and the diameter of the platform when the REM deprivation was manipulated. Some contradictory studies are considered, and it is shown that the discrepant outcomes in the literature could be the result of inattention to that ratio.

The purpose of this review is to focus attention on what appears to be a rather serious methodological oversight in the majority of the research that has measured the effects of depriving rats of REM sleep. The method generally employed is an adaptation of the platform-in-the-water technique first used by Jouvet and her colleagues (Jouvet, Vimont, Delorme, and Jouvet, 1964; Vimont-Vicary, Jouvet-Mounier, and Delorme, 1966). Using this technique, the deprivation of REM sleep is accomplished by maintaining the animal on a small platform surrounded by water. REM sleep is selectively eliminated because the loss of muscle tonus that accompanies the onset of the REM period causes the animal to contact the water, which either wakes the animal or causes it to return to non-REM sleep. The typical control for this procedure is to maintain the animal on a large platform, which reduces the likelihood the rat will contact the water at REM onset and hence diminishes the probability of the interruption of its REM sleep.

Jouvet's technique is compellingly simple and undoubtedly has contributed greatly to the recent proliferation of research on the matter. Generally, this technique has been used with rats and without any supporting electrophysiological data to assess the degree of deprivation of REM sleep actually achieved in the experimental and control conditions. Rather, the researchers involved seem to have been content to rely on data from a very limited literature (e.g., Duncan, Henry, Karadzic,

Mitchell, Pivik, Cohen, and Dement, 1967; Mendelson, Guthrie, Frederick, and Wyatt, 1974; Morden, Mitchell, and Dement, 1967), the studies which have validated Jouvet's technique with rats against the appropriate electrophysiological indices of sleep stages.

While the use of this technique with rats has been criticized because of its stressful nature (e.g., Stern, 1970; Stokes, 1973; Greenberg and Pearlman, 1974), a serious methodological difficulty, perhaps perpetuated by an uncritical acceptance of the forementioned validation studies, has been overlooked. The difficulty is outlined in Table 1. Shown there is a compilation of the diameters of the experimental (small) and control (large) platforms used in a number of the studies. If one compares these with the diameters of the platforms employed by Morden et al. (1967), the most commonly cited validation study, it is clear that there is little variation. That is, the diameters of the small and the large platforms have been relatively constant over this entire literature. However, as is also shown in Table 1, the size of the subjects, *when reported,* has not been constant but has ranged in weight from 130 to 455 grams and in age from 45 to 300 days.

Several observations may be derived from the studies summarized in Table 1. First, it is apparent that the size of the animals is a variable that has been generally considered unimportant. That is, studies typically report either incomplete or inexact information on the size of the animals and occasionally omit it completely. Second, with the exception of Mendelson et al. (1974), it seems that we have ignored the fact that the important determinant of the amount of deprivation of REM sleep with the small or large platform is not solely the diameter of the platform but rather the *ratio of the size of the animal to the diameter of the platform.*[1] Further, since the diameters of the platforms were relatively constant in the studies reviewed, it seems that the amount of deprivation achieved in this group of studies may have varied systematically as a function of the size of the animals used. Specifically, this is the hypothesis that the magnitude of the difference in deprivation between rats on small and large platforms varies inversely with the size of the animals. If this hypothesis is valid, one would expect the effects of deprivation more likely to be observed in studies using smaller (or younger) animals.

Validation of this hypothesis from the relevant literature is at best tenuous. Several factors seem to restrict the number of studies that can be used for this purpose. For example, a validation of our hypothesis requires both positive and negative results. Taken as a whole, the literature shows a positive bias. That is, only a few negative findings have been published. Also, both Holdstock and Verschoor (1973) and Greenberg

Table 1. The size (weight or age) of the animals and the diameter of the large and small platforms in studies of the effects of depriving rats of REM sleep

Study	Size		Diameter (cm)	
	Weight (grams)	Age (days)	Small platform	Large platform
Albert, Cicala, and Siegal (1970)	150–170	not given	7	11.5
Bowers, Hartmann, and Freedman (1966)	200–275	not given	5	not used
Boyaner (1970)	not given	not given	7	12
Hartmann and Stern (1972)	200–250	90	6	not used
Hicks, Johnson, and Sawrey (1972)	not given	45	6.25	13.32
Hicks and Paulus (1973)	not given	45	6.25	13.33
Hicks, Okuda, Pettey, and Thomsen (1976)	not given	45	6.25	13.33
Holdstock and Verschoor (1972)	230–270	77–105	6.5	not used
Holdstock and Verschoor (1973)	285–455	not given	6.5	11.5
Joy and Prinz (1969)	200	not given	7	14
Leconte and Bloch (1970a)	130–200	not given	7	12
Leconte and Bloch (1970b)	130–200	not given	7	12
Miller, Drew, and Schwartz (1971)	not given, not given	100–120, 300	7	14 5
Morden, Mitchell, and Dement (1967)	140–180, 260–280	45–48, 60–65	6.5	11.5
Morden, Conner, Mitchell, Dement, and Levine (1968)	220	90	7	11
Pearlman (1969)	not given	90–120	not given	not given
Pearlman (1971)	not given	90–120	7	not used
Pearlman (1973)	not given	90	7	not used
Pearlman and Becker (1973)	not given	120	7	not used
Pearlman and Greenberg (1973)	not given	90	7	not used
Sloan (1972)	284	80	7	11.5
Stern (1971a)	not given	65–90	5 5	11.5
Stern (1971b)	not given	not given	6	13

and Pearlman (1974) have argued that the different specific tasks used to define a given construct may vary considerably in the manner they are affected by the deprivation. Assuming that their arguments are valid, it is necessary to include only those studies that have used very similar procedures in studying the effects of the deprivation on a given variable. Otherwise, the predicted effects of inattention to the critical ratio (of the animal's size to the platform's diameter) could be obscured. Finally, a few studies that otherwise could have been used in this review were eliminated because the researchers simply failed to report the necessary data on the size of the animals. However, in spite of these restrictions, there

are enough studies in two areas of research — the studies of the effects of the deprivation of REM sleep on activity and the studies of its effects on acquisition and retention of conditioned avoidance — to provide some assessment of the validity of our hypothesis.

Four studies have measured the effects of the deprivation of REM sleep on the activity of rats, using a design that included both small and large platforms. Albert, Cicala, and Siegal (1970) found, using animals that weighed 150–175 grams, that the deprivation served to significantly enhance the level of activity. Judging from these weights, their animals were younger (and smaller) than those used by Stern (1971a) and Hicks, Thomsen, Pettey, and Okuda (1976). Stern used animals that were 65–90 days old and found, in contrast to Albert et al., no difference in activity between rats on small and large platforms. However, the rats in both of these groups were more active than those in a dry-cage control group. This finding could, in our frame of reference, be interpreted as demonstrating a positive relationship between amount of deprivation and level of activity. Hicks et al. (1976) used animals that were 45 days old and found no difference in activity between the rats in the groups on large and small platforms. They did, however, note that the animals in both groups should show significantly more activity than the rats in the control group with cold-water stress. Finally, Hicks, Okuda, Pettey, and Thomsen (1975) measured the effects of the deprivation of REM sleep on the level of activity of animals that were 30, 60, 90, and 120 days old. They partially controlled for the critical ratio of the size of the animal to the diameter of the platform in that the diameter of the small platform was altered for each age group. The diameter of the large platform was constant but somewhat larger, 16.5 cm, than has been typical (see Table 1). The probable result of this procedure was to reduce but not eliminate the effects of the ratio that is the focus of this paper. They found that the rats on small and large platforms for all age groups showed significantly more activity than the dry-cage controls. These data are similar to Stern's (1971a) and invite the same interpretation. In comparing the differences in level of activity between rats on small and large platforms for the different age groups, as might be expected, the youngest rats (30 days old) on the small platforms showed the strongest relationship between amount of deprivation and level of activity and replicated the relationship observed by Albert et al.

Two studies have measured the effects of the deprivation of REM sleep on the acquisition of an avoidance response, using a design that included both small and large platforms. Albert et al. (1970) used animals that weighed 150–175 grams. They reported a significant difference in the

rate of acquisition for the first 100 trials between the rats on small and large platforms. Using larger animals of 100–120 days old, Miller, Drew, and Schwartz (1971) found no difference between the groups on large and small platforms in acquisition of an avoidance response.

Three studies have measured the retention of conditioned avoidance after a period of deprivation of REM sleep. Pearlman (1969) found, in comparing his rats on small and large platforms, that retention of a passive avoidance response was disrupted to a greater degree in the rats on the small platforms. His animals were 90–120 days old. Leconte and Bloch (1970a, 1970b)[2] also reported that in comparison to the animals on the large platforms, those on the small platforms showed disrupted retention of a conditioned avoidance response. Judging from the weight range given, 130–200 grams, their animals were smaller than those used in Pearlman's study. In conflict with these studies, Miller et al. (1971) reported no difference in the retention of a passive avoidance response between the animals on small and large platforms. Their animals, which were 300 days old, were the largest we encountered in reviewing the literature. It is no surprise, then, to read that of their controls on the large platforms, "most . . . failed to take advantage of the greater surface area . . . while sleeping, and hence fell into the water almost as many times as" the animals on the small platforms.

To conclude, the results of this review support speculation by Stern (1970), Stokes (1973), and Greenberg and Pearlman (1974) that contradictions in the literature on the effects of depriving rats of REM sleep may, at least in part, be due to a lack of difference in deprivation between the conditions with small and large platforms. We have shown that these discrepancies may be the consequence of an uncritical use of Jouvet's technique, rather than some inherent defect of it. Clearly, the results of this review suggest that future research using Jouvet's technique should utilize apparatus that allows for consideration of the ratio of the size of the animal to the diameter of the platform. Further, the reporting of the data relevant to this ratio — the animals' ages and weights, and the platforms' diameters — should be made a conventional aspect of this literature. As Table 1 shows, the reporting of all of these data has been the exception rather than the rule.

Finally, and less directly, this review has called attention to a neglected variable in this area of research, namely, the age of the rats. Although we emphasized a methodological oversight in explaining certain contradictions in this literature, it is quite possible that, either partly or wholly, these discrepancies were the product of systematic differences of age that are inescapably associated with size. Age should be considered more care-

fully in future research, both as an experimental variable and as a potential source of confounding.

Notes

Dianne Thomsen is now at NASA–Ames Research Center, Moffett Field, California 94035.

1. Perhaps the most glaring example of inattention to this principle can be seen in the very study (i.e., Morden et al., 1967) that is usually cited as providing electrophysiological validation of Jouvet's technique for rats. Specifically, Morden and his colleagues used 45-day-old Sprague-Dawley rats on the small platforms and 60-day-old Long-Evans rats on the large platforms. That is, the size of the animals and the diameter of the platforms were confounded, presumably by design, in this the most commonly cited validation study.

2. We have referenced two papers by these authors. Although they bear different titles, the results are exactly the same, suggesting that the papers derived from the same set of data.

References

Albert, I., Cicala, G. S., and Siegal, J. 1970. The behavioral effects of REM-sleep deprivation in rats. *Psychophysiology* 6:550–560.

Bowers, M., Hartmann, E., and Freedman, D. 1966. Sleep deprivation and brain acetylcholine. *Science* 153:1416.

Boyaner, H. G. 1970. Effect of REM-sleep deprivation on exploration in rats. *Psychological Reports* 27:918.

Duncan, R., Henry, P., Karadzic, V., Mitchell, G., Pivik, T., Cohen, H., and Dement, W. 1967. Baseline sleep and REM-sleep deprivation in the rat. Paper read at the meeting of the Association for the Psychophysiological Study of Sleep in Santa Monica, California.

Greenberg, R., and Pearlman, C. 1974. Cutting the REM nerve: An approach to the adaptive role of REM sleep. *Perspectives in Biology and Medicine* 17:513–521.

Hartmann, E., and Stern, W. 1972. Desynchronized sleep deprivation: Learning deficit and its reversal by increased catecholamines. *Physiology and Behavior* 8:585–587.

Hicks, R. A., Johnson, J., and Sawrey, J. 1972. The effects of rapid-eye-movement sleep deprivation on ulceration in rats. Paper presented at the meeting of the Western Psychological Association in Portland, Oregon.

Hicks, R. A., Okuda, A., Pettey, B., and Thomsen, D. 1975. The effects of REM deprivation and age on exploration in rats. Paper presented at the meeting of the Western Psychological Association in Sacramento, California.

Hicks, R. A., and Paulus, M. 1973. Effects of rapid-eye-movement-sleep deprivation on the performance of rats in a T-maze. *Psychological Record* 23:89–92.

Hicks, R. A., Paulus, M., and Johnson, J. 1973. Effect of REM-sleep deprivation on electric-shock threshold in rats. *Psychological Reports* 32:1242.

Hicks, R. A., Thomsen, D., Pettey, B., and Okuda, A. 1976. REM-sleep deprivation and exploration in rats. *Psychological Reports* in press.

Holdstock, T. L., and Verschoor, G. J. 1972. Propensity for paradoxical sleep following deprivation of rats of different age groups. *Psychonomic Science* 29:39–40.

Holdstock, T. L., and Verschoor, G. J. 1973. Retention of maze learning following paradoxical sleep deprivation in rats. *Physiological Psychology* 1:29–32.

Jouvet, D., Vimont, P., Delorme, F., and Jouvet, M. 1964. Etude de la privation sélective de la phase paradoxale de sommeil chez le chat. *Comptes Rendues de la Société de Biologie* (Paris) 158:756–759.

Joy, R., and Prinz, P. 1969. The effect of sleep-altering environments upon the acquisition and retention of a conditioned-avoidance response in the rat. *Physiology and Behavior* 4:809–814.

Leconte, P., and Bloch, V. 1970a. Déficit de la rétention d'un conditionnement après privation de sommeil paradoxal chez le rat. *Comptes Rendus, Academie des Sciences* (Paris), series D, 271:226–229.

Leconte, P., and Bloch, V. 1970b. Effet de la privation de sommeil paradoxal sur l'acquisition et la rétention d'un conditionnement chez le rat. *Journal de Physiologie* (Paris) 62:290.

Mendelson, W. B., Guthrie, R. D., Frederick, G., and Wyatt, R. J. 1974. The flowerpot technique of rapid-eye-movement (REM) sleep deprivation. *Pharmacology, Biochemistry and Behavior* 2:553–556.

Miller, L., Drew, W., and Schwartz, I. 1971. Effect of REM-sleep deprivation on retention of a one-trial passive-avoidance response. *Perceptual and Motor Skills* 33:118.

Morden, B., Conner, R., Mitchell, G., Dement, W., and Levine, S. 1968. Effects of rapid-eye-movement (REM) sleep deprivation on shock-induced fighting. *Physiology and Behavior* 3:425–432.

Morden, B., Mitchell, G., and Dement, W. 1967. Selective REM-sleep deprivation and compensation phenomena in the rat. *Brain Research* 5:339–349.

Pearlman, C. 1969. Effect of rapid-eye-movement (dreaming) sleep deprivation on retention of avoidance learning in rats. U.S. Naval Submarine Medical Center, report 563.

Pearlman, C. 1971. Latent learning impaired by REM-sleep deprivation. *Psychonomic Science* 25:135–136.

Pearlman, C. 1973 REM-sleep deprivation impairs latent extinction in rats. *Physiology and Behavior* 11:233–237.

Pearlman, C., and Becker, M. 1973. Brief posttrial REM-sleep deprivation impairs discrimination learning in rats. *Physiological Psychology* 1:373–376.

Pearlman, C., and Greenberg, R. 1973. Posttrial REM sleep: A critical period for consolidation of shuttlebox avoidance. *Animal Learning and Behavior* 1:49–51.

Sloan, M. A. 1972. The effects of deprivation of rapid-eye-movement (REM) sleep on maze learning and aggression in the albino rat. *Journal of Psychiatric Research* 9:101–111.

Stern, W. C. 1970. The D-state, dreaming, and memory. *International Psychiatric Clinics* 7:249–257.

Stern, W. C. 1971a. Effects of desynchronized sleep deprivation upon startle-response habituation in the rat. *Psychonomic Science* 23:31–32.

Stern, W. C. 1971b. Acquisition impairments following rapid-eye-movement-sleep deprivation in rats. *Physiology and Behavior* 7:345–352.

Stokes, J. P. 1973. The effects of rapid-eye-movement sleep on retention. *Psychological Record* 23:521–531.

Vimont-Vicary, P., Jouvet-Mounier, D., and Delorme, F. 1966. Effets EEG et comportementaux des privations de sommeil paradoxal chez le chat. *Electroencephalography and Clinical Neurophysiology* 20:439–449.

AMERICAN JOURNAL OF PSYCHOLOGY
March 1977, Vol. 90, No. 1, pp. 103–119

An expectancy model for estimating the effects of age and incentive in probability learning

Jack R. Nation and John C. McCullers
Texas A&M University and University of Oklahoma

A theoretical model for estimating the effects of age and incentive on children's performance in probability learning is introduced. The preliminary considerations and assumptions of the model are presented, along with a U-shaped curve predicted by the model. The implications of the model and its predictions about performance by different age groups given different incentives are discussed, and it is shown that the model integrates apparently contradictory findings in the literature. Novel predictions of the model about the effects of combinations of age and incentive were tested and confirmed.

There are some diverse, and seemingly contradictory, findings on the probability learning of children. While numerous factors would be expected to influence these subjects' decision-making processes, a comprehensive review of the literature suggests that age and incentive are the major determinants of responding in probability learning and, in that respect, potentially responsible for the apparently conflicting results. The present paper introduces an empirically derived model to account for the discrepant findings in the literature and, more important, tests several novel predictions of this model to confirm its validity and usefulness as a tool for research.

THE MODEL PROPOSED

Goodnow (1955) suggested that one of the determinants of performance in probability learning is the level of success the subject is willing to accept. For example, if the subject expects a high degree of success, a perfect solution, then he does *not* adopt a maximal-gain strategy,[1] because he is unwilling to overlook an occasional loss. Rather, he searches for a perfect solution to the problem, with the result that his response frequencies tend to match the relative frequencies of the task stimuli: probability matching. Probability matching relative to a maximal-gain

strategy leads to a decrease in the number of correct (reinforced) choices. Thus, the higher the subject's expectancy of achieving a perfect solution, the greater the likelihood that he will not choose a maximal-gain strategy (e.g., Stevenson and Zigler, 1958; Stevenson and Weir, 1959; Gruen, Ottinger, and Zigler, 1970; Gruen and Zigler, 1968). On the other hand, apart from differences in response strategies related to the subject's interest in finding a perfect solution, there is a tendency to adopt a maximal-gain strategy with rewards of higher incentive value (Brackbill, Kappy, and Starr, 1962; Offenbach, 1964).

In order to adequately estimate a subject's overall expectancy of success, it would seem useful to distinguish between the initial expectancy held by the subject at the beginning of the experiment and the expectancy that is the result of what happens during the experiment. The former expectancy might be thought of as a 'subject-determined' expectancy, influenced by age, cultural experiences (socioeconomic status, education, child-rearing practices of the parents, etc.), personality characteristics (birth order, anxiety level, need for achievement, conformity, etc.), sex, intellectual status (gifted, retarded), recent reinforcement history (in the school, home, etc.), level of academic achievement, and so on. The latter expectancy would be more a 'task-determined' factor associated with the experimental situation itself, influenced by variables such as incentive value, reinforcement schedules, instructions, amount and type of pre-training, social factors (sex of experimenter, social reinforcement, etc.), type of task, apparatus, number of response alternatives, the contingency of reward and punishment, and so on. Thus, a basic assumption of the present model was that performance is determined by an overall 'expectancy of success,' E, that is a joint function of a subject-determined expectancy, e_S, and a task-determined expectancy, e_T.[2]

Subject-determined expectancy, e_S

In order to keep the model arithmetically simple, it was further assumed that an individual's e_S for a given situation varies along a continuum from 0 to 1 as a function of past experience and various subject factors. Although age was the principle determinant of e_S for the purpose of the present paper, it was assumed that an individual's e_S could also be influenced by other variables such as sex, intelligence, socioeconomic status, and personality characteristics.

It was assumed that younger children have cognitive, intellectual, and motor skills inferior to those of older children and would therefore be less successful in dealing with complex daily problems. Simply, the more

facile the individual, the more likely he or she would be to achieve given goal. Thus, the e_S of younger children would be lower than th of older children.

Task-determined expectancy, e_T

It was assumed that rewards vary along a continuum from 0 to 1 i terms of their incentive value. Just as e_S may be influenced by a variet of subject factors, it was assumed that e_T could be influenced by a variet of task factors. For simplicitly, e_T was restricted primarily to incentiv value here.

Effective overall expectancy, E

It was assumed that e_S and e_T combine in a multiplicative manne This assumption emerged as a result of empirical curve fitting. Other as sumptions (additive, etc.) were attempted, but none provided a curv that fit the data as well as the multiplicative assumption.[3] The mod thus involves two factors, e_S and e_T, that combine to produce an effectiv overall expectancy of success, E, such that

$$e_S \cdot e_T = E. \qquad \text{[Equation}$$

Although E was assumed to be an increasing function of combinatior of e_S and e_T, it was not necessary to assume that performance is an ir creasing function of E. Because of the different strategies employed wit different overall expectancies, performance could be nonmonotonicall related to E. Specifically, it was assumed that at low levels of E, perfor mance would be high because the subject is willing to overlook a lo (Stevenson and Weir, 1959); that at middle levels of E, performanc would be low because the subject is no longer willing to overlook a lo and thus resorts to matching (Stevenson and Hoving, 1964); and th at high levels of E, performance would again be high because there is s much for the subject to win by using a maximal-gain strategy (Brackbi et al., 1962). The result is that performance in probability learning shoul be a U-shaped function of E.

Such a function was observed by Weir (1964) in a development analysis of probability learning. Weir (1964) combined data from sever studies using a three-choice task to produce a picture of development changes in probability learning among subjects from 3 to 20 years of ag All subjects were tested under quite similar experimental conditions, an Weir (1964) noted that the subjects in every age group seemed to b

highly motivated. Weir's results on terminal response levels yielded a U-shaped function. In terms of the present model, Weir (1964) held task-determined expectancy (e_T) constant and varied subject-determined expectancy (e_S) by manipulating age. As mentioned previously, factors other than age are considered to contribute to e_S, but they would be expected to have been random in Weir's (1964) study. Weir's (1964) results with the three-choice task have been replicated with a two-choice task (Derks and Paclisanu, 1967; Sullivan and Ross, 1970).

A CURVE CONSTRUCTED

The curve shown in Figure 1 was determined by an empirical curve-fitting effort that made use of data from several published studies of probability learning by children. It has been assumed that e_S and e_T combine to generate a U-shaped function. While Figure 1 represents a preliminary approximation to the hypothesized shape, a precise mathematical specification of the function has not been attempted. Nevertheless, Figure 1 allows theoretical levels of effective expectancy to be established that identify the points at which low performance should occur ($E = .07$) and at which high performance again would be expected ($E = .17$). At present, the only predictions that can be made are in terms of two broad categories of measurement: high versus low performance. Table 1 summarizes the results of the ten studies of probability learning that manipulated age and incentive and were used to construct the curve shown in Figure 1.

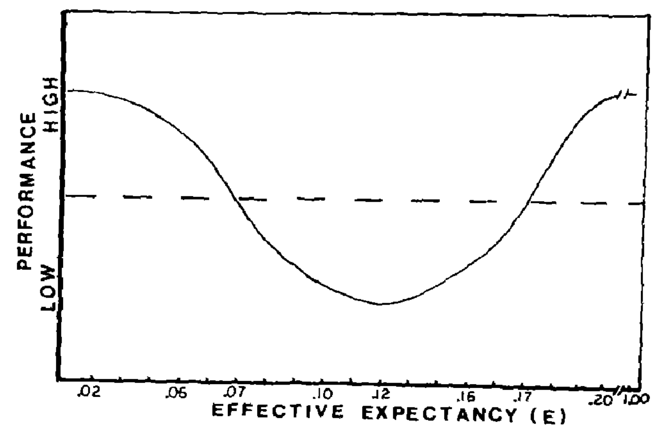


Figure 1. U-shaped function of E used to predict performance in probability learning. The effective expectancies .07 and .17 represent the two theoretical points on the curve that separate high and low performance

Table 1. Predictions of the model and obtained performance

Subjects (and incentive value of the rewards)	e_S ·	e_T =	E	Performance: Predicted	Performance: Obtained[a]
Stevenson and Zigler, 1958					
Mentally retarded (middle)	.05	.6	03	high	82%
Normal five-year-olds (middle)	2	.6	.12	low	65%
Stevenson and Weir, 1959					
Five-year-olds (low)	.2	.2	.04	high	85%
Five-year-olds (middle)	.2	.6	.12	low	75%
Brackbill, Kappy, and Starr, 1962					
Nine-year-olds (low)	.8	.2	.16	low	.62[b]
Nine-year-olds (middle)	.8	.6	.48	high	.72[b]
Offenbach, 1964					
Nine-year-olds (low)	.8	.2	.16	low	68[b]
Nine-year-olds (middle)	.8	.6	.48	high	79[b]
Stevenson and Hoving, 1964					
Five-year-olds (low)	.2	.2	04	high	80%
Five-year-olds (middle)	2	.6	12	low	61%
Nine-year-olds (low)	.8	2	16	low	65%
Nine-year-olds (middle)	.8	.6	48	high	78%
Adults (low)	1.0	2	20	high	81%
Adults (middle)	1.0	.6	60	high	79%
Weir, 1964					
Three-and-a-half-year-olds (middle, 33% reinforcement)	.05	6	03	high	70%[c]
Five-year-olds (middle, 33% reinforcement)	2	6	12	low	59%[c]
Eighteen-year-olds (middle, 33% reinforcement)	1 0	6	60	high	67%[c]
Bisett and Rieber, 1966					
Six- and seven-year-olds (low, subject-determined)	.5	.15	08	low	49%
Six- and seven-year-olds (high, subject-determined)	.5	9	.45	high	62%

[a] Mean percent correct unless otherwise noted. When the exact results were not available, an estimate was obtained from the graphs provided by the respective articles. Although the data obtained may appear to have been based on different dependent variables, in all cases the dependent variable was in fact proportional or percent choice of the more frequent event (or more frequently reinforced stimulus). Because of different experimental conditions across studies, however, comparisons should be restricted to differences within studies.

[b] Mean proportional choice of most probable stimulus.

[c] Mean percent correct on last block of 20 trials.

Table 1 continued

Subjects (and incentive value of the rewards)	e_S ·	e_T =	E	Performance Predicted	Obtained[a]
Ten-year-olds (low, subject-determined)	.8	.15	.13	low	52%
Ten-year-olds (low, subject-determined)	.8	.9	.72	high	58%
Derks and Paclisanu, 1967					
Three-and-a-half-year-olds (middle)	.05	.6	.02	high	.87[d]
Five-year-olds (middle)	.2	.6	.12	low	.62[d]
Twenty-one-year-olds (middle)	1.0	.6	.60	high	.77[d]
Gruen and Zigler, 1968					
Middle-class (middle)	.2	.6	.12	low	65%
Lower-class (middle)	.05	.6	.03	high	80%
Mentally retarded (middle)	.05	.6	.03	high	80%
Sullivan and Ross, 1970					
Five-year-olds (low)	.2	.2	.04	high	80%
Nine-year-olds (low)	.8	.2	.16	low	72%
Twelve-year-olds (low)	1.0	.2	.20	high	81%

[d] Proportional prediction of more frequent event.

That is, those ten studies provided the data base by which the curve in Figure 1 was conceived. The curve was specifically detailed to fit these data and therefore has no independent existence apart from the results listed in Table 1. The studies listed in Table 1 were selected to reflect the discrepant findings on age and incentive that appear in the literature and to include some of the more frequently cited studies on this topic. As may be seen in Table 1, the specific e_S and e_T employed remain constant across studies for given developmental levels and incentive values. Notice also the unusually low levels of e_S (.05) for retarded children, three-and-a-half-year-olds, and children of lower socioeconomic status. While these low levels were obtained by a curve-fitting procedure, they appear to represent the lower end of the e_S continuum. This suggests that whenever the functioning of young children is impaired by either intellectual or cultural deficiency, e_S drops to a level that parallels that of the ordinary three-and-a-half-year-old.

In summarizing the model, attention should be directed toward the interrelationship of the published data that appear in Table 1 and the curve derived from these data that appears in Figure 1. The authors by no means wish to imply that the predictions of the results in Table 1 are

novel; indeed, they are not. Quite the reverse is the case, since the spe values of e_S and e_T were determined by these data. Considering thi is not surprising that all of the prediction rules listed in Table 1 hold the present model. It would be strange if the rules did not apply, s they were in fact determined by the very data they predict. The achi ment in fitting the curve in Figure 1 to the data in Table 1 is simply such a curve *can* be specified. This specification, hopefully, will assis accounting for the discrepant results often reported in studies of dren's probability learning.

THE LITERATURE INTEGRATED

As noted from Table 1 and in the previous discussion, one accomp ment of the present model is to integrate a fairly large literature that h tofore appeared to be in conflict. It seems worthwhile to examine, acc ing to prediction rules permitted by Equation 1, several of the experim involved in the derivation of the model. It should again be noted th reality, the model was derived from these data and therefore non the predictions is a priori. What the model does achieve, however, meaningful account of some apparently contradictory results.

Several investigators (e.g., Stevenson and Hoving, 1964; Stevenson Weir, 1959) have observed high performance by five-year-olds wit wards of low incentive value. According to the model, the logic i follows. Because of motor and cognitive differences, preschoolers are as likely as older children to be successful problem solvers, and there the *five*-year-olds' initial expectancy (e_S) was assumed to be low, .2 a reward of *low* incentive value ($e_T = .2$) is provided, the result f Equation 1 would be

(five-, low) $.2 \cdot .2 = .04.$ (high performa

In Figure 1 it can be seen that an effective expectancy of .04 woul expected to yield *high performance*. Thus, the model successfully acco for high performance by five-year-olds with rewards of low incer value.

A similar analysis can be made of the prediction rule generated older children. The older subjects' greater success with solvable probl and workable tasks should result in a higher initial expectancy than of five-year-olds, and e_S was set at .8 for the nine-year-olds. If a rev of low incentive value is used with these older children, then, effe expectancy would become

(nine-, low) $.8 \cdot .2 = .16.$ (low performa

Thus, the prediction is that nine-year-olds would be inferior in performance to five-year-olds when the incentive value of the reinforcer is low. Such results have been obtained empirically by Stevenson and Weir (1959) and were used in constructing the curve shown in Figure 1.

The model predicts essentially equal performance for younger children with rewards of middle incentive value ($e_T = .6$) and older children with rewards of low incentive value ($e_T = .2$). These values are

(five-, middle)	$.2 \cdot .6 = .12$, and	(low performance)
(nine-, low)	$.8 \cdot .2 = .16$.	(low performance)

Thus, the prediction is that performance should be low for both groups. This is an interesting prediction, and inspection of the graphs of Stevenson and Hoving (1964) reveals that the performance of five-year-olds with rewards of middle (high) incentive value essentially paralleled the performance of nine-year-olds with rewards of low incentive value.[5]

There have been several reports (Brackbill et al., 1962; Offenbach, 1964; Stevenson and Hoving, 1964) that older children maximize their gain with rewards of high incentive value; that is, the greater the reward, the more likely these subjects are to use a strategy that results in a high-reinforcement payoff. In terms of the present model, the effective expectancy for the older subjects ($e_S = .8$) with rewards of a middle incentive value ($e_T = .6$) is rather high,

(nine-, middle)	$.8 \cdot .6 = .48$.	(high performance)

It can be seen in Figure 1 that .48 is well above the critical level (.17) of E required to generate high performance. Therefore, the higher performance of older children with rewards of middle (high) incentive value (e.g., Brackbill et al., 1962; Offenbach, 1964) is consistent with the model and represents another example of the efficacy of the curve in Figure 1.

With rewards of low incentive value, the model yields this prediction for older children:

(nine-, low)	$.8 \cdot .2 = .16$.	(low performance)

An effective expectancy of .16, as may be seen in Figure 1, leads to a prediction of poorer performance. This prediction is compatible with the finding that older children perform better with rewards of middle (high) incentive value than with those of low (Brackbill et al., 1962; Offenbach, 1964).

There is one study that may at first appear to be inconsistent with the model being proposed here. Bisett and Rieber (1966) found that both younger children and older children perform better with rewards of high

incentive value when that value was determined by the subjects. However, the younger children in their study were not the usual five-year-olds, but *six-* and *seven*-year-olds. The older children were *ten*-year-olds, rather than nine-year-olds. School experiences, age, and other factors would be expected to result in some increase in the value of e_S both for six- and seven-year-olds and for ten-year-olds. If it is assumed that e_S for six- and seven-year-olds is .5 (approximately half the distance between the e_S for five-year-olds, .2, and the e_S for nine-year-olds, .8)[6] and that the e_S for ten-year-olds is .85 (slightly higher than the .80 assigned to nine-year-olds), then the prediction equations for rewards of low and high incentive values become

(six- and seven-, low)	$.5 \cdot .15 = .08$,	(low performance)
(six- and seven-, high)	$.5 \cdot .9 = .45$,	(high performance)
(ten-, low)	$.85 \cdot .15 = .13$, and	(low performance)
(ten-, high)	$.85 \cdot .9 = .77$.	(high performance)

Note that e_T for rewards of a high incentive value is .9. The reason is that the values of Bisett and Rieber's incentives were *subject*-determined rather than experimenter-determined.[7] The use of rewards of empirically determined incentive values ensures that the subjects actually receive rewards that have those values. When each subject receives a reward that is of high subject-determined incentive value, e_T more closely approximates the maximal value of 1.0. Similarly, the e_T with rewards of low subject-determined incentive value decreases from .20 to .15. Figure 1 reveals that with these effective expectancies, no differences in performance would be expected between the younger and older children. The results of Bisett and Rieber (1966) are then consistent with predictions of the model.

One of the most intriguing predictions of the model is that there should be no differences in adults' performance regardless of the differences in the incentive values of the rewards. The reason is that even with rewards of low value, the e_S of adults is so high (at 1.0) that almost any level of e_T will yield an E that is asymptotic (i.e., at or above .17 in Figure 1). Note in Table 1 that this prediction is supported by the findings of Stevenson and Hoving (1964).

THE MODEL TESTED

The model assumes that performance is a joint function of the age of the subject and the incentive value of the reward. Thus, one implication of the model is that the performance of subjects at given ages can be altered by manipulating incentive value. Specifically, the model predicts that five-year-olds will do well with rewards of low incentive value, do

more poorly with rewards of middle value, and finally, do well again with rewards of high value. The effective expectancy of a younger subject with rewards of high incentive value should be functionally equivalent to that of an older child with rewards of high value. The rewards of middle incentive value are necessary to verify the general shape of the function being proposed; further, they lead to differential predictions for younger and older children, predictions not tested in previous studies. From Equation 1, the predictions for younger children (five-year-olds) and older children (nine-year-olds) with rewards of low, middle, and high incentive value are

(five-, low)	$.2 \cdot .15 = .03$,	(high performance)
(five-, middle)	$.2 \cdot .6 = .12$,	(low performance)
(five-, high)	$.2 \cdot .9 = .18$,	(high performance)
(nine-, low)	$.8 \cdot .15 = .12$,	(low performance)
(nine-, middle)	$.8 \cdot .6 = .48$, and	(high performance)
(nine-, high)	$.8 \cdot .9 = .72$.	(high performance)

Thus, the purpose of the present experiment was to test the model with rewards at three levels of incentive value and two age groups.

METHOD

Subjects

The 30 five-year-olds used as subjects (17 boys and 13 girls) were enrolled in the nursery school and laboratory school at the Child Development Center of the University of Oklahoma. The 30 nine-year-olds (15 boys and 15 girls) were enrolled in a public elementary school. The sample was relatively homogeneous in terms of race and socioeconomic status, as the area is primarily white and middle-class.

Experimental material

The experimental material consisted of 3-by-5-in. white index cards. On each card was placed a picture of an apple or an orange. Two decks of ten cards each were prepared in order to counterbalance the apples and oranges. The relative proportion of more frequent to less frequent events was held constant for all treatments at 7:3. The additional materials were the rewards: chocolate kisses, marbles, or paper clips.

Design and procedure

A 2 (ages) × 3 (incentive values) factorial design was used in the study. The specific kind of reward given to each subject was determined individually according to his assigned condition for incentive value and his rating of each type of reward as of high, middle, or low incentive value. The subjects participated individually.

Each subject was instructed at the outset on the "rules of the game." The

deck was placed face down and the subject's task was to guess whether a picture of an apple or an orange would appear when the experimenter turned the card over. For each correct prediction, the subject received a reward appropriate to the condition for incentive value to which he had been assigned. No rewards were given for incorrect predictions. Each subject was given 80 trials. The experimenter reshuffled the deck of ten cards after every trial. Thus, the probabilities remained the same for any given trial. The rewards the subject won were placed in a plastic bag and held until the end of the school day, at which time the subject could pick up his earnings.

RESULTS

Incentive value

The ratings of the rewards as of high, middle, or low incentive value are presented in Table 2. While it can be seen that the five-year-olds were more variable than the nine-year-olds, there was a high degree of agreement about the relative incentive value of the three kinds of rewards. That is, for the majority of subjects, the candy had a high incentive value, the marble a middle value, and the paper clip a low value.

Probability learning

Since the majority of studies cited in Table 1 examined terminal probability responding, separate analyses were performed on the first 40 trials and the last 40 trials of the present experiment in order to facilitate theoretical discussion. A 2 (ages) × 3 (incentive values) × 4 (blocks of trials) analysis of variance was performed on the mean numbers of choices of the more probable stimulus over the *first* 40 trials. None of the main effects or tests for interaction reached an acceptable level of significance, although the interaction of age by incentive value did approach significance [$F(6, 162) = 1.91, p < .10$]. So although preliminary changes did begin to occur during the first 40 trials of the experiment, the differences were not statistically reliable.

Table 2. Mean percent choices of each type of reward as of high, middle, or low incentive value

	Five-year-olds			Nine-year-olds		
	High	Middle	Low	High	Middle	Low
Candy	70.00	23 30	6 70	86 67	13.33	
Marble	26 67	70.00	3 30	13 33	86 67	
Paper clip	3.30	6 70	90 00			100.00

The mean numbers of choices of the more probable stimulus over the *last* 40 trials are presented in Figure 2 in blocks of 10 trials. As may be seen there, performance was essentially at chance for the five-year-olds with rewards of middle incentive value and for the nine-year-olds with rewards of low value. All other combinations of age and incentive value yielded performance that approximated probability matching. It may be noted that the younger children's predictions of the more probable stimulus were about 5% above, while the older children's predictions fell about 5% below, the objective probability of occurrence (i.e., 70%). While this slight difference in performance between the two age groups is consistent in both magnitude and direction with that reported by several other investigators, it is not a prediction of the model proposed here. That is, at this time the model is designed to estimate performance quite broadly, as high or low, without attempting to distinguish gradations of performance within these two categories.

A 2 (ages) × 3 (incentive values) × 4 (blocks of trials) analysis of variance was performed on the data over the last 40 trials. The main effect of blocks was nonsignificant [$F < 1.00$], indicating that performance was asymptotic over the last 40 trials. In addition, all interactions with trials were nonsignificant [all Fs < 1.59]. The effect of interest is the interaction of age by incentive value. The analysis revealed this interaction to be highly significant [$F(2, 54) = 20.59$, $p < .0001$]. Tukey's post hoc comparisons indicated that for the five-year-olds, performance with rewards of middle incentive value was significantly inferior [$p < .01$] to that with rewards of both low and high incentive value, which latter did not differ [$p > .05$]. For the nine-year-olds, performance with rewards

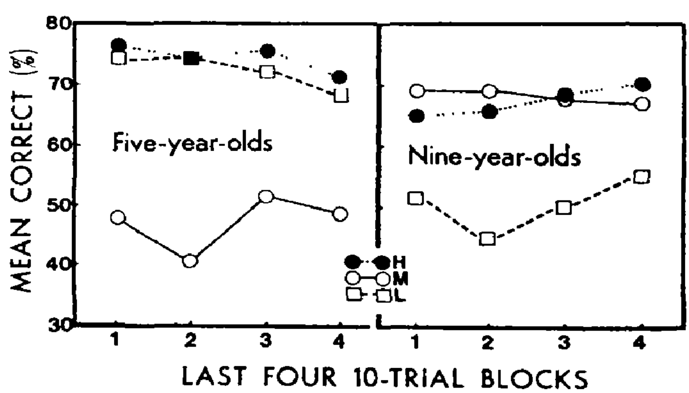


Figure 2. Mean percent correct (choices of the more probable stimulus event) across the last four 10-trial blocks for five- and nine-year-olds given rewards at three levels of incentive value (high, middle, and low)

of both middle and high incentive value was superior to that with rewards of low incentive value [$p < .01$], which former did not significantly differ [$p > .05$].

Patterned responding

Although many patterns of responding may have occurred, the frequencies with which subjects employed single alternations or double alternations were selected for analysis. Separate 2 (ages) × 3 (incentive values) × 8 (blocks of trials) analyses of variance were performed. For single alternations, none of the main effects or tests for interaction was statistically significant, although differences with level of incentive value did marginally occur [$F(2, 54) = 2.45, p < .10$]. These differences were primarily a result of a decrease in patterned responding for both five-year-olds and nine-year-olds with rewards of high incentive value. For double alternations, the only statistically significant result was the main effect of age, indicating more frequent patterned responding by nine-year-olds than by five-year-olds [$F(1, 54) = 39.19, p < .001$]. All other tests failed to approach significance [all Fs < 1.20].

DISCUSSION

The present results are consistent with those of Stevenson and Weir (1959) and Stevenson and Hoving (1964) in that five-year-olds performed better with rewards of low incentive value than with rewards of a middle value. However, it was additionally shown that when rewards of high incentive value were provided, the performance of five-year-olds was again high, closely approximating their performance with rewards of low value. The present results are also in accord with the findings of other studies in probability learning (Brackbill et al., 1962; Offenbach, 1964; Stevenson and Hoving, 1964) in showing that older children performed better with rewards of high incentive value than with those of low value. One additional finding of the present investigation was that the older children with rewards of middle and high incentive value showed no differences in performance. This verifies a prediction of the model, since E in both cases fell above the critical .17 shown in Figure 1.

The present data provide an interesting corollary to Weir's (1964) finding of a curvilinear relation between frequency of correct response and age. Weir (1964) showed that performance changes as a function of e_S when e_T is held constant. The present data indicate that the converse is also true — that performance changes as a function of e_T when e_S is

held constant. Weir's (1964) data coupled with those of the present investigation support the notion of a U-shaped function that is sensitive to manipulations of both age (e_S) and incentive (e_T). Together, these data supply evidence that both task and subject variables, as outlined in the present report, are potent determinants of children's performance in probability learning.

The analyses of the subjects' tendencies to develop patterned responding revealed that the nine-year-olds consistently used double alternations more often than the five-year-olds. These data may reflect a fundamental difference between the two age groups in ability to develop sequentially dependent cognitive strategies, but they suggest no underlying strategy that would account for the changes in performance found in the present experiment. The complex method by which children of different ages process information remains obscure.

SUMMARY

The present article should be perceived as a general theoretical description of children's performance in probability learning. Perhaps any model devised to explain the data in this literature would have to be couched in very general terms, at least initially. However, with further refinement, the present model could perhaps become a mathematically precise predictive tool to be used scientifically by future investigators of probability learning. This precision can be most likely achieved by extending the boundary conditions of the model and investigating the influence of other factors, presently unspecified.

The extension of the model to include special populations of children would certainly contribute toward a more elaborate predictive system. For example, Table 1 provides a preliminary indication of how the model might be adapted to questions of intellectual (Stevenson and Zigler, 1958) and socioeconomic differences (Gruen and Zigler, 1968). While the roles of such factors as emotional stability and personality are less clear, they doubtless are involved in the formulation of individual subjects' strategies. Whatever the variables regulating children's performance in probability learning, exact specification of e_S and e_T can only come by empirical examination in well-controlled experiments. Simply, the more data there are to involve in the curve-fitting enterprise, the more accurate will be the model.

It is noteworthy that, potentially, the present model could serve as a device for measuring the value of an incentive, given that a subject's expectancy is known. Specifically, given that E and e_S are known for any

individual, then one need only solve for the unknown e_T to obtain an index of incentive value. The efficacy of such a procedure, of course, depends on how exactly E and e_S can be specified. Thus, the ability to measure incentive value within the framework of the model will increase with increasing quantification of parameters of the model.

The incorporation of individual differences into the present model may be useful at a somewhat more applied level. In working with normal children in the classroom, the teacher may be able to use information about how a child of a given age develops strategies for problem solving with incentives of a given value. That is, by scrutinizing specific patterns of responding on a probability-learning task, the educator may be better able to assess the cognitive skills of a child he or she has in the classroom. It would be interesting to compare the patterns of responding of different groups of children such as the mentally retarded, intellectually gifted, emotionally disturbed, and so on, and determine to what extent these children use similar and dissimilar approaches to problem solving. Such information might prove useful in designing educational programs.

In conclusion, this paper has attempted to resolve a long-standing discrepancy in the literature on probability learning and to provide a theoretical model that might serve as a guide for future investigation. The results of this preliminary effort appear encouraging. One accomplishment was to provide a reasonable indication of the manner in which the model is to be applied in probability-learning situations, as the thrust of a theory is communicated as much by example or by application to concrete experimental settings as by abstract understanding of the several principles that comprise the theory. Yet, much work remains to develop and specify its exact parameters. Only through future research will the genuine worth of the model be determined. The need is for more data.

Notes

The authors express their gratitude to Stephen S. Roop and Richard W. Dickinson for their assistance in the conduct and analysis of this experiment. Received for publication October 25, 1975; revision, January 12, 1976.

1. A maximal-gain strategy is that whereby the subject consistently responds to the more frequently reinforced event.

2. At this point, 'expectancy' is merely a convenient working label. It implies no commitment to any existing theoretical orientation. Its choice as a working label was based partly on its frequent use by other investigators working with children's probability learning. While other theoretical terms (e.g., incentive motivation) carry similar meaning, expectancy seems better suited to the present

analysis for two reasons. First, it appears to be more appropriate to the strategies for pattern solving apparently involved in children's learning than those stimulus/response terms that have emerged from the animal laboratory. Second, it is more compatible with the molar level of conceptualization employed in this paper.

3. Elements of the present model are similar in many ways to existing theories (e.g., those of Hull and Spence, Rotter's theory of social learning). However, those theories as originally proposed do not include developmental parameters. Developmental considerations appear necessary in order to account for the unusual and contrary-to-logic findings that in some cases younger children perform better than older children (Stevenson and Weir, 1959; Stevenson and Hoving, 1964). Nevertheless, readers who prefer to conceptualize the present effort in terms of existing theories should not find it difficult to render a workable translation.

4. The values for e_S and e_T are empirical estimates derived by fitting published data to the curve in Figure 1. Even though the specific estimated values are applied consistently throughout the paper, they represent no more than preliminary approximations. It may be possible at some time to quantify e_S and e_T to the extent that exact values can be specified.

5. Incentive value in a number of studies to be considered here was estimated by the experimenter rather than scaled by the subject, and the same kinds of rewards were not given consistent values from study to study. Accordingly, the labeling of a specific type of reward as of high, medium, or low incentive value was based on the reported values if those values were determined empirically; otherwise, the incentive values were estimated from studies in which a number of types of rewards had been scaled by the subjects (e.g., Bisett and Rieber, 1966; McCullers and Martin, 1971). Where the estimation of incentive value for present purposes differs from that reported by the author(s), the value as originally reported is shown in parentheses in the text; it does not appear in Table 1.

6. If e_S were assumed to be a simple linear function of age, the interpolated e_S for six- and seven-year-olds combined would be .425. However, it is logical to assume that subject-determined expectancy climbs at a more rapid rate at higher points on the age continuum (i.e., the growth function is nonlinear). Thus, the e_S for six- and seven-year-olds is set at .5.

7. When incentive values are experimenter-determined, the e_T assigned across subjects migrates toward the center of the continuum, toward .5. When incentive values are subject-determined, the actual e_T scores for subjects within a group become more homogeneous. Because of this homogeneity of variance, the distribution is not likely to be skewed and thus mean e_T will be higher (or lower) in studies where the determination of incentive value as high (or low) is subject-determined as opposed to experimenter-determined. Note that this difference occurs only when rewards of high or low incentive value are being used. With rewards of a middle value, there is no floor or ceiling effect operating to produce a skewed distribution. The effect of using rewards of a middle value as determined by the experimenter would be to increase the variability of the individual e_T scores within a group of subjects about the group mean, but the mean e_T across subjects would remain the same whether the incentive values were experimenter-determined or subject-determined.

References

Bisett, B. M., and Rieber, M. 1966. The effects of age and incentive value on discrimination learning. *Journal of Experimental Child Psychology* 3:199–206.

Brackbill, Y., Kappy, M. S., and Starr, R. H. 1962. Magnitude of reward and probability learning. *Journal of Experimental Psychology* 73:278–285.

Derks, P. L., and Paclisanu, M. 1967. Simple strategies and binary predictions by children and adults. *Journal of Experimental Psychology* 73:278–285.

Goodnow, J. J. 1955. Determinants of choice distribution in two-choice situations. *American Journal of Psychology* 68:106–116.

Gruen, G., Ottinger, D., and Zigler, E. 1970. Level of aspiration and the probability learning of middle- and lower-class children. *Developmental Psychology* 3:133–142.

Gruen, G., and Zigler, E. 1968. Expectancy of success and the probability learning of middle-class, lower-class, and retarded children. *Journal of Abnormal Psychology* 73:343–352.

McCullers, J. C., and Martin, J. A. G. 1971. A reexamination of the role of incentives in children's discrimination learning. *Child Development* 42:827–837.

Offenbach, S. I. 1964. Studies of children's probability-learning behavior: I. Effect of reward and punishment at two age levels. *Child Development* 35:709–715.

Stevenson, H. W., and Hoving, K. L. 1964. Probability learning as a function of age and incentive. *Journal of Experimental Child Psychology* 1:64–70.

Stevenson, H. W., and Weir, M. W. 1959. Variables affecting children's performance in a probability-learning task. *Journal of Experimental Psychology* 57:403–412.

Stevenson, H. W., and Zigler, E. 1958. Probability learning in children. *Journal of Experimental Psychology* 56:185–192.

Sullivan, F. J., and Ross, B. M. 1970. What is learned in probability learning. *Developmental Psychology* 2:58–65.

Weir, M. W. 1964. Developmental changes in problem-solving strategies. *Psychological Review* 71:473–490.

ANNOUNCEMENT

Midwestern Association of Behavior Analysis: Annual Meeting

The Midwestern Association of Behavior Analysis is an interdisciplinary group of people interested in applied and experimental analysis of behavior. On May 1–4, 1976, the Association held its Second Annual Meeting at the Blackstone Hotel in Chicago. Indications of its success include an increased attendance over the previous year, with 1,250 attending; a large variety of papers, symposia, workshops, and invited addresses; and many special events, including a Testimonial Dinner for Dr. B. F. Skinner and Dr. Fred Keller.

The Third Annual Meeting is to be held May 14–17, 1977, again at the Blackstone Hotel in Chicago. If you wish to know more about the organization and its past or planned activities, please write or call:

Linda Parrott
Corresponding Secretary
Midwestern Association of Behavior Analysis
Department of Psychology
Western Michigan University
Kalamazoo, Michigan 49008
616/383-1629

AMERICAN JOURNAL OF PSYCHOLOGY
March 1977, Vol. 90, No. 1, pp. 121–131

The effect of blank trials on probability learning

C. Weir
University College London

In an experiment on the effects of blank trials on probability learning, some informational parameters were varied. The results showed that the presence of blank trials shifted response probabilities toward the guessing level. The data from other experiments are considered, and the relevance of the results to studies of behavior with concurrent schedules is discussed.

This paper reports an experiment on the effect of blank trials on probability learning. A blank trial (*Eo*) is when neither reinforcing event (*E*1 or *E*2) is presented after a prediction response (*A*1 or *A*2) is made. In this paper, only the noncontingent case of probability learning will be considered, where reinforcing events do not depend on responses.

The experiment to be reported here can be considered to be of a more general interest than with probability learning per se. The data may be applicable to investigations of learning under almost any partial-reinforcement regime, especially that with concurrent schedules where subjects get different reinforcement probabilities on each schedule. Blank trials are characteristic of such schedules; for instance, the concurrent schedules VI-1 and VI-3 provide reinforcement after every minute (on average) on one schedule and after every three minutes on the other. Many responses on such schedules go unreinforced.

Both in probability learning and with concurrent schedules, a *matching* of response probabilities to reinforcement probabilities is typically obtained. Matching has been of interest to operant researchers recently. Herrnstein (1974) and Baum (1974) have specified some theoretical and logical properties of matching. In 1974, Baum argued that matching will not be obtained when the discrimination between alternative responses is poor. His arguments may be applicable to probability learning, and there may already be some direct tests of them in the literature on probability learning. Clearly, then, there would be an advantage gained if a close connection between matching in probability learning and with concurrent schedules could be established.

A problem with such a parallel is that there is a contingency of rein forcement on response enforced with concurrent schedules but not ir probability learning. This issue will be unresolved here, although it i: difficult to separate the two situations from the subject's point of view: in probability learning, the subject usually interprets a confirmed prediction as though a contingency had been in force. Most of the recent 'runs' theories to describe the subject's behavior in probability learning make this assumption explicit (see Jones, 1971).

Previous attempts to study the effect of blank trials on probability learning have given rise to somewhat discrepant conclusions. Atkinson (1956) and Neimark (1956) concluded that blank trials had no effect on matching, while Anderson and Grant (1957, 1958), Laberge, Greeno, and Peterson (1962), and Greeno (1962) found some effects. The main finding of the latter group of authors was that blank trials shift the probability of response 1, $p(A1)$, nearer to the .50 guessing level and, therefore, further from the matching level, $p(E1)$ or π_1. This effect will be referred to as undermatching, using Baum's terminology. A conventional stimulus-sampling model of probability learning assumes the null effect of blank trials, giving weight to the results of Atkinson and Neimark (see Atkinson, Bower, and Crothers, 1965, pp. 343–372).

My experiment was performed to determine whether apparently discrepant results in the literature on probability learning can be understood in terms of the availability of the *informative stimuli*, $E1$ and $E2$. When blank, Eo, trials are inserted into probability learning, the informative stimuli necessarily occur less frequently. Undermatching could result when blank trials are present because the average number of informative stimuli per minute, their rate of presentation, is decreased. Alternatively, undermatching could result simply because the total number of $E1$ and $E2$ trials is insufficient to enable the subject to assess $p(E1)$ and hence match $p(A1)$ to $p(E1)$. Either situation could result in poor discrimination of the differential reinforcement of $A1$ and $A2$. Some data that Baum mentions would predict undermatching with blank trials.

METHOD

There were three groups of subjects, all receiving 120 $E1$ and $E2$ trials, three $E1$ trials for every $E2$ trial, $p(E1/E1 \text{ or } E2) = \pi_1 = .75$. Two groups received no blank trials but served as controls. The experimental group (group X) received 80 blank, Eo, trials, for a total of 200 trials.

For the first control group (group Y), trial duration was longer than for the experimental group. It was chosen so that the average rate of presentation of informative stimuli, E_1 or E_2, was the same for the two groups. If undermatching

occurs when blank trials are present because of a slower average rate of presentation of the informative stimuli, then this control group would have a depressed $p(A1)$ similar to that of the group with blank trials. Results of studies where this rate was varied in probability learning are not in complete agreement, although generally, no systematic effect has been found (Grant, Hornseth, and Hake, 1950; Lakota and Madison, 1971).

The second control group (group Z) had the same total number of informative trials at the same trial duration as the experimental group. If $p(A1)$ for this control group was like that for the experimental group, then undermatching might be due simply to the scarcity of $E1$ and $E2$ events when blank trials are present. Table 1 presents the relationships between all three groups on the important parameters.

Subjects

There were 32 subjects, assigned randomly to one of the three groups. There were 16 in group X and 8 each in groups Y and Z. Most of these subjects were students at University College London, while the others were of varied occupation. Their ages ranged from 18 to 26 years, with a median of 19. No subject was paid for serving in this experiment.

Apparatus

Each subject was seated at a table approximately 50 cm wide. He looked toward a sheet-metal panel, 127 by 186 mm, on which two buttons and three lights were mounted. The lights were central, and the middle one, which served as a warning signal, was a different color than the outer ones. The two outer 6-V flashlight bulbs were used as reinforcing stimuli, $E1$ and $E2$. The two buttons, placed directly underneath the outer lights, were for the responses, $A1$ and $A2$. The side where the most frequent light occurred was designated $E1$. Every 20 trials, the number of $A1$ responses was recorded manually, as these had been cumulating on an index counter.

Procedure

In group X, the subjects were presented 200 noncontingent trials for probability learning, with three $E1$ trials for every $E2$ trial. Interspersed with the $E1$ and $E2$ trials were Eo (blank) trials, so that neither $E1$ nor $E2$ occurred on 40% of the trials. For every signal, the subjects were required to press button $A1$ or $A2$; and $E1$ or $E2$ or Eo was presented according to a predetermined

Table 1. The three groups, their probabilities of reinforcing events, rate of presentation of informative trials, and trial duration

	$p(E1)$	$p(Eo)$	$p(E1/E1$ or $E2)$	Average rate of $E1$ and $E2$	Trial duration
Group X	45	40	75	6/min	6 sec
Group Y	75	0	.75	6/min	10 sec
Group Z	.75	0	.75	10/min	6 sec

random sequence. Three different orders of reinforcing events were used in balanced fashion. Each sequence was randomized completely in 20 trials: *I* for 9 trials, *E*2 for 3 trials, *E*o for 8 trials. For half the group, *E*1 was the rig light; for the other half, it was the left light. The button beneath the *E*1 lig was designated *A*1.

Trial duration in group X was fixed at 6 sec (see Table 1). The followi timing was used: a 2-sec signal, an .8-sec interstimulus interval, a 2-sec rei forcing event, a 1.2-sec intertrial interval. The experimental session lasted 20 mi

The experimenter began by reading typical instructions for probability lear ing (Estes and Straughan, 1954) to each subject. No mention was made blank trials. The subject was asked to pose questions during the practice tria but not during the experimental session. The practice trials had the outcom *E*1, *E*2, *E*1, *E*2, *E*2, *E*1.

In group Y, subjects were presented the 120 *E*1 and *E*2 trials that group received, but no blank trials were included. The trial duration was increased 10 sec so that the experimental session for this group lasted 20 min. In this wa the informative trials (*E*1 and *E*2) were presented to subjects in group Y the same average rate as to subjects in group X: six informative trials p minute. The trials had a 2.4-sec interstimulus interval and a 3.6-sec intertri interval. This group was the control for rate of informative stimuli.

In group Z, again, no blanks were presented, although the *E*1 and *E*2 tria were given. The trial duration and sequences of informative trials were the san as for group X, but the experimental session only lasted 12 min. Group Z w the control for total number of informative stimuli per se.

In short, all groups had $p(E1/E1 \text{ or } E2) = \pi_1 = .75$. Group X received 40 blank trials. Groups Y and Z received no blank trials, just the same 120 info mative trials given group X. They differed from each other in rate of present tion and trial duration.

RESULTS

Terminal responding

Figure 1 presents the main findings for groups X, Y, and Z. The mo striking aspect of the graph is that the group with blank trials had a lowe $p(A1)$ at the end of the experiment than the groups with no blank tria [$p(A1) = .61$ versus .76, $SE = .04$ and .03].

Taking the last half of the sessions, thus equating for informative trial analysis of variance showed a significant effect of blank trials [$F(1, 28) = 9.14, p < .005$]. No difference was found between groups Y and Z, th two groups with no blank trials [$F(1, 28) = 2.38, p < .1$], nor betwee the subgroups of group X [$F(1, 28) < 1$].[1] The comparison betwee groups Y and Z also tested for a reliable effect of trial duration. The sin ple effect of blank trials emerging from these tests confirms the finding of Laberge et al. (1962), Greeno (1962), and Anderson and Grar (1957).

There was a difference between $p(A1)$ in the first and second halve

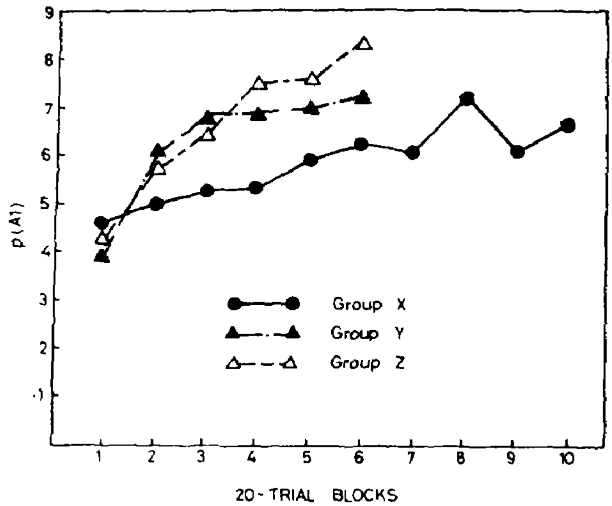


Figure 1. Average $p(A1)$ over 20-trial blocks for group X (blank trials) and groups Y and Z (no blank trials)

of the session [$F(1, 30) = 42.8, p < .001$], as compared to a lack of difference between the fifth and sixth (informative) trial blocks [$F < 1$]. This finding does not contradict the assumption that the subjects were at asymptote; however, it does not prove that the subjects were at a stable level even though the standard error was only about .03 at the end of the session. Unfortunately, there is no critical test to determine whether asymptotic responding was attained; but at least differences over trial blocks were diminishing.

Tests for other effects were performed. The sequence of $E1$ and $E2$ did *not* significantly affect $p(A1)$ [$F(2, 24) = 1.14$], despite recent discoveries of such effects (see Jones, 1971). Whether $E1$ was assigned to the right or left light on the panel was not a reliable factor affecting $p(A1)$ either. These tests were performed on the last 40 trials, a less conservative measure than for the other statistics.

Rate of learning

Using the first six 20-trial blocks, the parameter θ was calculated for each group by the technique of Estes and Straughan (1954). These estimates were much lower for the group with blank trials than for the other two groups, being .007 for group X as compared to .033 and .049 for groups Y and Z respectively. This finding indicates that Eo affects $p(A1)$ by depressing the rate of learning.

DISCUSSION

The undermatching with blank trials that has been reported in the literature on probability learning can thus be attributed to a dilution of the $E1$ and $E2$ trials by the Eo trials, since group X differed from groups Y and Z. This dilution probably makes differences between $A1$ and $A2$ less salient to subjects. Since groups X and Y had an equal average rate of presentation of informative stimuli, this information-processing variable is not critical. Similarly, the difference between groups X and Z showed that the total number of informative stimuli is not itself the critical variable.

This effect of blank trials is due to a difference in rate of learning, at least, and the question about their effect at asymptote is still an open one. If blank trials have such a profound effect on matching in probability learning, then perhaps the parallel between behavior in probability learning and with concurrent schedules is only superficial.

The discrepancies in the literature

Since the experiment confirmed the trend toward undermatching, a closer scrutiny of the literature seems in order. A graph of the matching ratio, $p(A1)/p(E1)$, over levels of $p(Eo)$ was prepared for the relevant studies. If the ratio is unity, matching of $p(A1)$ to π_1 is obtained; if less than unity, undermatching occurred. Figure 2 shows the data from studies by Anderson and Grant, Greeno, Laberge et al., Neimark, Weir (1970; from the dissertation, but not the experiment, reported here), and the experiment reported in this paper. An indirect relation between matching and $p(Eo)$ is readily noticeable. The 15 points in the figure represent independent groups of subjects who were presented two-choice trials for probability learning, with trials of 5 to 6 sec in duration and blank trials interspersed in a noncontingent manner. The regression line based on 15 groups, $p(A1)/p(E1) = -.24p(Eo) + 1.0$, accounts for 46% of the variance in the matching ratio. Considered as a whole, then, there is considerable evidence for undermatching as a function of $p(Eo)$.

In light of this finding, another look at the data of Neimark (1956) and Atkinson (1956) is in order, since they concluded that Eo had no effect on $p(A1)$. Neimark used relatively low $p(Eo)$, .34 and .17, and a relatively short sequence of trials, 100. Perhaps if she had increased $p(Eo)$ to .6 or .7, her conclusion would have been different. The regression line shows that her data followed the trend of the other studies plotted.

Atkinson used a procedure with Eo contingent on $A1$ responses. This

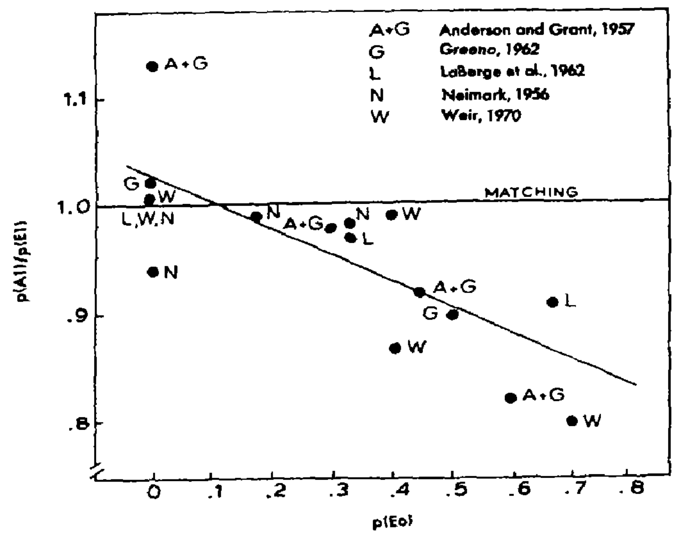


Figure 2. Matching ratios as a function of $p(Eo)$ for 15 independent groups of subjects from five studies of probability learning

procedure may have been noticed by his subjects and makes his experiment difficult to compare with the rest (hence it was omitted from the regression analysis). The data on contingent reinforcement suggest that all such effects are complex and differ from those with noncontingent reinforcement (Borkowski and Johns, 1966). Also suspect is Atkinson's conclusion that the model postulating no effect of Eo is the best one of those tested. The prediction from this 'identity' model yielded a range of $p(A1)$ levels, while his other model predicted specific $p(A1)$ levels. Therefore, he was more likely to accept the identity model, assuming $p(A1)$ is a random variable. A fairer test of the models with all predictions equally specific will be needed before his conclusion is generally acceptable. Nevertheless, if Atkinson had not used contingent reinforcement, his results would debase the relationship found in the data from the six other experiments. Leaving open the question of the importance of contingent reinforcement, then, there is a clear empirical relationship between $p(A1)$ and $p(Eo)$.

Theoretical considerations

For theories of why blank trials affect probability learning, one can turn either to the models of probability learning or to explanations of behavior with concurrent schedules. If we assume that probability learn-

ing with blank trials is analogous to the situation involving concurrent schedules, we must postulate that undermatching is obtained for the same reasons in both settings. (Note that there are some methodological differences between the paradigms.) In operant terms, the problem is to specify why poor discrimination of *A*1 and *A*2 schedules is increased with π_0. That is, we must find the common elements between the case with blank trials and the three cases of undermatching cited by Baum (1974).

Since the models of probability learning are more formal, it may be easier to modify them. The stimulus-sampling model postulates that blank trials have no effect (based on Neimark, and Atkinson): the so-called identity model. So a possible modification in the light of the evidence would be that a certain proportion, b, of blank trials will have the effect of an incorrect prediction response (and on $1 - b$ trials, no effect). The consistent evidence of an effect of blank trials seems to justify this addition.

Such a model would predict a mean learning curve resulting in matching when b or π_0 is zero. If π_0 exceeds zero, then undermatching would be predicted. This, of course, is consistent with most of the evidence cited in Figure 2 (Weir, 1970). However, if this model is used to predict sequential dependencies (Weir, 1970, Experiment 3), then it is no more successful than the conventional stimulus-sampling model. Because predicting both sequential dependencies *and* mean learning curves is difficult for most models, recent interest has concentrated on sequences (Jones, 1971), although none of the recent theories has addressed itself to probability learning with blank trials.

In general, then, despite inadequacies, a model accounting for blank trials is at least as good as the identity model in describing sequential dependencies and is better than the identity model in describing undermatching of mean learning curves. This better fit is not surprising since the blank-trials model has one additional parameter, but it is disappointing that sequential dependencies were not described better.

Operant models of behavior based on trial-by-trial operations (Shimp, 1969; Killeen, 1972; Rachlin, 1973) have looked primarily at data on the behavior of pigeons and have worked out various trial-by-trial mechanisms whereby a pigeon changes its response pattern. Generally, the schemes suggested to control behavior have been based on the local rate of reinforcement in the particular situation studied. Although there may be a spurious relationship between the rate of blank trials and the local rate of reinforcement, the choice of a maximally reinforcing response would be the same no matter how many blank trials. So these models do not directly add to our understanding of the data reported here.

Instead, Rachlin's (1973) discussion of economic and biological prin-

ples operating on choice responding seems more relevant. The economic rinciple works in the long run, while the biological principle works in ie short run. The latter is similar to Shimp and Killeen's momentary aximizing of reinforcement, while the obtained effect of blank trials on ıman probability learning seems to be relatively permanent. One might terpret these findings as showing that information about the availability ' *reinforcement* is diluted by blank trials and affects the subject's assess- ent of the reinforcement value of the alternative responses. This assess- ent would be part of the operation of Rachlin's economic principle or milar to Baum's discriminability theory.

Assuming subjects are governed by an economic principle, they could › comparing the relative reinforcement rate or cost of the responses, $A1$ ıd $A2$. If one calculates the reinforcement rates and costs involved in ie studies discussed in this paper, the following observations can be made. he difference between $A1$ and $A2$ in terms of reinforcement rate de- eases as $p(Eo)$ increases. In the extreme case, there would be no dif- rence between reinforcement rates of $A1$ and $A2$ when $p(Eo)$ is 1.0. hus, subjects would have more difficulty distinguishing reinforcement tes at high π_o values. If reinforcement rate is the parameter the subjects e monitoring, then $p(E1)$ would be closer to $p(E2)$ and more under- atching would result as a function of π_o.

On the other hand, if reinforcement cost (number of responses per inforcement) is the critical parameter, the reinforcement costs for $A1$ ıd $A2$ would each increase as $p(Eo)$ increases. Undermatching could sult only if the subject was selectively sensitive to large changes in rein- rcement cost. Then he might discriminate the additional cost of $E1$ but ›t $E2$, and his $p(A1)$ would be reduced to reflect this selective bias. nce the qualification that subjects notice only large changes would be quired to argue that reinforcement cost is the controlling variable, it ould seem simpler to conclude the effect of blank trials found here sup- ›rts the notion that subjects are monitoring reinforcement rates of the ternative responses.

To conclude, the present experiment showed undermatching when ank trials were interspersed among the trials for probability learning. he literature considered as a whole strongly supports this finding; how- er, a simple modification of the stimulus-sampling model was disap- ›inting. The data are interpreted to support a mechanism by which bjects evaluate the relative reinforcement rate of $A1$ and $A2$ to deter- ine their responding, a procedure which may result in poor discrimi- ıbility between schedules.

Notes

This work was supported by the Marshall Aid Commemoration Commission and is part of a doctoral dissertation done at the University of London (see Weir, 1970). The author thanks Professor R. J. Audley and Dr. S. Supramaniam for their help in editing this paper. Received for publication September 25, 1975; revision, January 12, 1976.

1. Half of the subjects in group X and those in groups Y and Z were given a slightly modified task for reasons irrelevant to the question of the effect of blank trials. They were given $E1$ or $E2$ contingent on their responses on 10% of the trials. Since the mean $p(E1/E1 \text{ or } E2)$ was .77 ± .02 and their response patterns were essentially the same, the data of these subjects in group X were added to those of the more conventional group X.

References

Anderson, N. H. 1964. An evaluation of stimulus sampling theory. In *Categories of human learning*, ed. A. W. Melton. New York: Academic Press.

Anderson, N. H., and Grant, D. A. 1957. A test of a statistical learning theory model for two-choice behavior with double stimulus events. *Journal of Experimental Psychology* 54:305–317.

Anderson, N. H., and Grant, D. A. 1958. A correction and reanalysis. *Journal of Experimental Psychology* 56:453–454.

Atkinson, R. C. 1956. An analysis of the effect of nonreinforced trials in terms of statistical learning theory. *Journal of Experimental Psychology* 52:28–32.

Atkinson, R. C., Bower, G. H., and Crothers, E. J. 1965. *An introduction to mathematical learning theory*. New York: Wiley.

Baum, W. M. 1974. On two types of deviation from the matching law: Bias and undermatching. *Journal of the Experimental Analysis of Behavior* 22:231–242.

Baum, W. M. 1975. Time allocation in human vigilance. *Journal of the Experimental Analysis of Behavior* 23:45–53.

Borkowski, J. G., and Johns, G. R. 1966. Analysis of learning rate and sampling probabilities in a contingent-reinforcement situation. *Journal of Experimental Psychology* 72:158–160.

Estes, W. K., and Straughan, J. H. 1954. Analysis of a verbal-conditioning situation in terms of statistical learning theory. *Journal of Experimental Psychology* 47:225–234.

Grant, D. A., Hornseth, J. P., and Hake, H. W. 1950. The influence of intertrial interval on the Humphreys 'random reinforcement' effect during the extinction of a verbal response. *Journal of Experimental Psychology* 40:609–612.

Greeno, J. G. 1962. Effects of nonreinforced trials in two-choice learning with noncontingent reinforcement. *Journal of Experimental Psychology* 64:373–379.

Herrnstein, R. J. 1974. Formal properties of the matching law. *Journal of the Experimental Analysis of Behavior* 21:159–164.

Jones, M. R. 1971. From probability learning to sequential processing: A critical review. *Psychological Bulletin* 76:153–185.

Killeen, Peter. 1972. A yoked-chamber comparison of concurrent and multiple schedules. *Journal of the Experimental Analysis of Behavior* 18:13–22.

Laberge, D., Greeno, J. G., and Peterson, O. 1962. Nonreinforcement and neutralization of stimuli. *Journal of Experimental Psychology* 63:207–213.

Lakota, R. A., and Madison, H. L. 1971. Luminance and reinforcement delay in probability learning. *Journal of Experimental Psychology* 88:277–278.

Neimark, E. D. 1956. Effects of type of nonreinforcement and number of alternative responses in two verbal-conditioning situations. *Journal of Experimental Psychology* 52:209–220.

Rachlin, Howard. 1973. Contrast and matching. *Psychological Review* 80:217–234.

Shimp, C. P. 1969. Optimal behavior in free operant experiments. *Psychological Review* 76:97–112.

Weir, C. J. 1970. An analysis of a two-choice guessing situation terms of concurrent-reinforcement schedules. Doctoral dissertation, University of London.

ANNOUNCEMENT

Institute for Social Research: Survey Research Center summer institute

For the thirtieth consecutive year, the Survey Research Center of the Institute for Social Research at the University of Michigan will hold a summer institute in survey research techniques from June 27 to August 20, 1977.

The Institute is designed to meet some of the educational and training needs of men and women engaged in business, government research, and other statistical work, and also to meet the needs of graduate students and university instructors interested in quantitative research in the social sciences. For further information on specific dates, courses, fees, or housing, write

Helene J. Hitchcock
Administrative Manager, Office of the Director
Survey Research Center, Institute for Social Research
P.O. Box 1248
Ann Arbor, Michigan 48106

AMERICAN JOURNAL OF PSYCHOLOGY
March 1977, Vol. 90, No. 1, pp. 133–144

Individual differences related to performance on two word-recognition tasks

Daniel M. Platnick and Larry G. Richards
University of Virginia

Tests of cognitive and perceptual abilities were given to 140 college students. These subjects then participated in two word-recognition tasks in which their recognition thresholds to both English and artificial words were obtained. Analysis revealed two components (an 'overall' and a 'familiarity' component) underlying the recognition thresholds and showed that of the five hypothesized test factors, speed of closure related to the overall component while perceptual speed and memory span related to the familiarity component.

Two of the best established facts about word recognition are that visual recognition thresholds vary inversely with word frequency both for English words (Howes and Solomon, 1951) and for artificial words (Solomon and Postman, 1952). These two findings are both called the *word-frequency effect* and are usually assumed to reflect the same underlying processes, although Richards and Platnick (1974a, 1974b) have argued that they represent different processes.

Two general classes of theories were initially proposed to explain these findings. Osgood (1964) maintained that the word-frequency effect is a *perceptual* phenomenon. Frequent patterned stimulation is a sufficient condition for the formation of sensory integrations. A subject sees a frequently encountered word more easily and more clearly than a seldom encountered word because he has a sensory integration (perceptual unit) corresponding to the word, a unit whose strength is a function of the number of his previous experiences with the word. On the other hand, various investigators (Goldiamond and Hawkins, 1958; Richards, 1973) have asserted that the effect is a phenomenon of response bias or *memory*. In terms of Solomon and Postman's paradigm, this model involves both demand characteristics and memory of the pretraining list. In terms of Howes and Solomon's paradigm, the model involves 'sophisticated guessing' (Neisser, 1967).

One strategy for exploring the processes involved in an experimental task is to see how differences between individuals relate to performance on the task (Cronbach, 1957). In the present context, one can explore whether perceptual or memory factors are important by determining whether performance in word recognition correlates with performance on tests of perceptual or memory abilities.

Spielberger and Denny (1963) selected two groups of subjects differing in verbal ability. Recognition thresholds were obtained tachistoscopically for English words of low, moderate, and high frequency. The only group differences found were for low-frequency words: subjects of high verbal ability had lower thresholds for low-frequency words than did subjects of low verbal ability. One possible interpretation is that subjects of high verbal ability have had more frequent exposure to low-frequency words. Another interpretation is that there are differences in perceptual or cognitive abilities between subjects of high and low verbal ability.

Richards and Platnick (1974a) combined the approach of Spielberger and Denny with the experimental tasks of Howes and Solomon and of Solomon and Postman. They selected 80 subjects, 40 with verbal scores of 700 or greater and 40 with verbal scores of 500 or less on the CEEB Scholastic Aptitude test. Subjects were exposed to Solomon and Postman's artificial words during a pretraining session; these words occurred at various frequencies (1, 5, 10, 25). Recognition thresholds were then obtained for English words of low, medium, and high frequency, after which they were obtained for the artificial words. Word frequency significantly influenced recognition thresholds for both English and artificial words. No main effect of verbal ability was found in either situation. However, for the English words, a highly significant interaction of verbal ability and frequency was found, replicating the earlier finding of Spielberger and Denny. No such interaction was found for the artificial words: for these words, the curves for subjects of high and low verbal ability were essentially the same. This suggests that Spielberger and Denny's result was not a matter of different levels of perceptual organization by these two groups of subjects but rather of their differential experience with English words. When the two groups have the same experience, as they did in the training with artificial words, the interaction of verbal ability and frequency vanishes.

The present paper reports an extension of this type of study to other measures of perceptual and cognitive abilities. Tests were selected to assess factors on perceptual and cognitive abilities presumably involved in word recognition. Subjects were given these tests, after which they were in-

volved in two word-recognition tasks. Performance on the word-recognition tasks was then related to the outcome of the tests.

METHOD

Subjects

The subjects were 140 volunteers from introductory psychology classes at the University of Virginia. All subjects had normal or corrected-to-normal visual acuity, and all had English as their native language.

Tests

Tests from the kit of tests for cognitive factors (French, Ekstrom, and Price, 1963) were given to all subjects in the first part of the study. These tests were selected as indices of several factors assumed to influence word recognition.

One of these factors, *word fluency,* is a facility in "producing isolated words that contain one or more structural restrictions without reference to the meaning of words" (French et al., 1963). Modified versions of the Word Endings and Word Beginnings tests were used as measures of word fluency. In order to decrease loadings on verbal ability (Guilford and Hoepfner, 1971) for both of these tests, only one-letter, rather than multiple-letter, specifications were given to the subjects.

Another factor, *memory span,* is the ability "to recall perfectly for immediate reproduction a series of items after only one presentation of the series" (French et al., 1963). The Visual and Auditory Number Span tests were used as measures of memory span. For the auditory test, the numbers were presented orally by the experimenter. The 24 sequences presented ranged in length from four to twelve digits. For the visual test, the numbers were projected sequentially on a screen at the front of the room using Kodak Carousel projectors. Four long sequences were deleted from this test, so that only 20 sequences were presented, from four to eleven digits in length.

Yet another factor, *perceptual speed,* is defined as a facility in "finding figures, making comparisons, and carrying out other very simple tasks involving visual perception" (French et al., 1963). The Number Comparison and Identical Pictures tests were used to assess perceptual speed.

The final factor here, *speed of closure,* is the ability to "unify an apparently disparate perceptual field into a single percept" (French et al., 1963). It was assessed using the Gestalt Completion and Concealed Words tests.

There were two testing sessions. Subjects initially completed a general information form that asked their verbal and quantitative SAT scores, whether English was their native language, and whether they needed corrective lenses.

The tests were administered in the same order during both testing sessions. One of the two tests on each factor appeared in the first half of a session, and the other test on the same factor appeared in the latter half of the session. The other restriction imposed was that the two tests on the same factor be separated by at least two intervening tests. The order of administration of the tests was Number Comparison, Word Beginnings, Gestalt Completion, Auditory Number, Word Endings, Identical Pictures, Concealed Words, and Visual Number. The

instructions suggested by French et al. were used throughout. Both the Auditory Number and the Visual Number tests were each administered as a whole, whereas each of the other tests was administered in two independently timed subparts.

Thresholds

Two weeks after the testing sessions, subjects returned for the second part of the experiment. At this time, they were given word-recognition tasks to perform. Each subject initially underwent pretraining on a set of artificial words. The artificial words were taken from the list of Solomon and Postman (1952). All were seven-letter three-syllable words. A pack of 92 3-by-5-in. index cards was presented to the subject. Each card contained one of the pronounceable artificial words: experimental words, which occurred at various frequencies (0, 1, 5, 15 times); and filler words, each of which occurred only once. Four different decks of artificial words were used in order to counterbalance the experimental words over frequencies. The deck was randomly shuffled for each new subject. The subject examined each card and read aloud its word.

Then, using a tachistoscope, duration thresholds were obtained for the 12 English words, and immediately afterward, for the 16 artificial words. The English words also had seven letters and three syllables each. They were selected from the Thorndike-Lorge word count (1944) to represent three levels of word frequency (low, 1–10 per million; medium, 20–30 per million; and high, A and AA).

Thresholds were obtained using a Scientific Prototype two-channel tachistoscope (model 800F). Both the stimulus field and the blank field were set at luminances of 32 mlam, as measured by a Macbeth illuminometer. The blank field contained two parallel lines to guide the subject's fixation to that part of the visual field where the stimulus words would appear. The subject initiated the stimulus presentation by using a hand trigger. Stimulus duration was varied in determining thresholds. A set of four preliminary words was used to familiarize the subject with the tachistoscope and to obtain an estimate of his thresholds for high-frequency words. The ascending method of limits was used for determining the threshold of each word, starting each sequence at 10 msec and increasing duration in steps of 5 msec until a duration of 70 msec was attained, and in steps of 10 msec from 70 to 200 msec. The series was terminated when the subject either correctly identified the stimulus two consecutive times or had gone 200 msec without correctly identifying it. For both English and artificial words, the order of presentation was counterbalanced over levels of frequency.

RESULTS

Tests

With the exception of the Gestalt Completion and Identical Pictures tests, the tests were at an appropriate level of difficulty for the subjects. The Gestalt Completion and Identical Pictures tests lacked an adequate ceiling: many subjects attained perfect scores on them. Means on the two subparts of the test on Word Endings differed significantly, and that

test had the lowest reliability, .59. The letter M proved to be an extremely difficult word-ending cue to use. For both of the Word Beginnings subparts and for one of the Word Endings subparts (words ending with N), high-frequency words were emitted before low-frequency ones, as one would expect from the 'spew principle' (Kintsch, 1970). Spearman-Brown reliabilities for all tests except the test on Word Endings were in the range from .70 to .87.

The correlations between the ten tests were obtained. The correlations between the two tests measuring any given factor were found to be higher than the correlations of any tests measuring different factors. A principal-components analysis was performed on this correlation matrix. Since the tests had been selected as measures of four factors of cognitive ability, at least four components were hypothesized, plus a fifth for the SAT scores. Eigenvalues greater than unity are associated with the first five components, and those five components each account for a relatively large proportion of the total variance. Components beyond the fifth were discarded. The first five principal components account for 77.5% of the total variance.[1]

These five components were rotated, using Varimax, to an approximation of simple structure. The resulting matrix of factor loadings appears in Table 1. The hypothesized factor structure is clearly evident. Factor 1 is *word fluency,* Factor 2 is *memory span.* Factor 3 is *speed of closure:* in addition to the Gestalt Completion and Concealed Words tests, a moderate loading is evident for Identical Pictures. Factor 4 has equal, large loadings from the two SAT tests and can be considered to reflect *intellectual performance.* Factor 5 is *perceptual speed.* Thus, for this sample of subjects, the hypothesized factor structure was clearly evident. What will be called derived (factor) test scores were computed for each subject on each of the five rotated components.

Thresholds

Figure 1 shows mean recognition thresholds on the tachistoscopic tasks for both the artificial and English words. The English words were recognized faster than the artificial words, and both sets of words displayed an inverse relation between word frequency and recognition thresholds. For English words, there was no real difference in mean threshold for the medium- and high-frequency words, but the drop from low- to medium-frequency words was quite large.

The sum of the recognition thresholds for the four words of each type and frequency was obtained for each subject. With these sums as data,

Table 1. Varimax rotated factor matrix for test scores

Test	Factor				
	1 Word fluency	2 Memory span	3 Speed of closure	4 Intellectual performance	5 Perceptual speed
Auditory Number Span	03173	.89439	.06330	.03261	.09679
Visual Number Span	.19904	.82526	07971	.12202	−.10629
Word Endings	.89418	.09719	.15289	.05596	−.03851
Word Beginnings	.85697	13390	.02237	.09493	.23527
Gestalt Completion	.09723	−.03240	.86896	− 01923	.04909
Concealed Words	05759	22567	.75741	.20608	−.00009
Number Comparison	.01663	09613	− 13022	02894	.89462
Identical Pictures	.23567	− 16517	.39420	05891	.71184
Verbal SAT	16476	−.01769	03790	.85366	.07937
Quantitative SAT	− 02259	16717	11937	84249	− 00938

correlations were obtained between words of different types and frequencies taken across subjects. The resulting matrix is shown in Table 2. The correlations between words of the same type, English or artificial, tend to be higher than those between words of different types.

A principal-components analysis was done on the correlation matrix in Table 2. The first principal component, component I, accounts for about 80% of the variance in thresholds. All measures load highly on this component: it is a general task component, an average or *overall* threshold component. The second component, component II, accounts for 9% of the variance. Figure 2 shows that the recognition thresholds to English words load negatively on this component, while the recognition thresholds to artificial words load positively. Low-frequency artificial words cluster at one end of component II, and higher-frequency English words cluster at the other; high-frequency artificial and low-frequency English words fall in between. Thus, this second threshold component

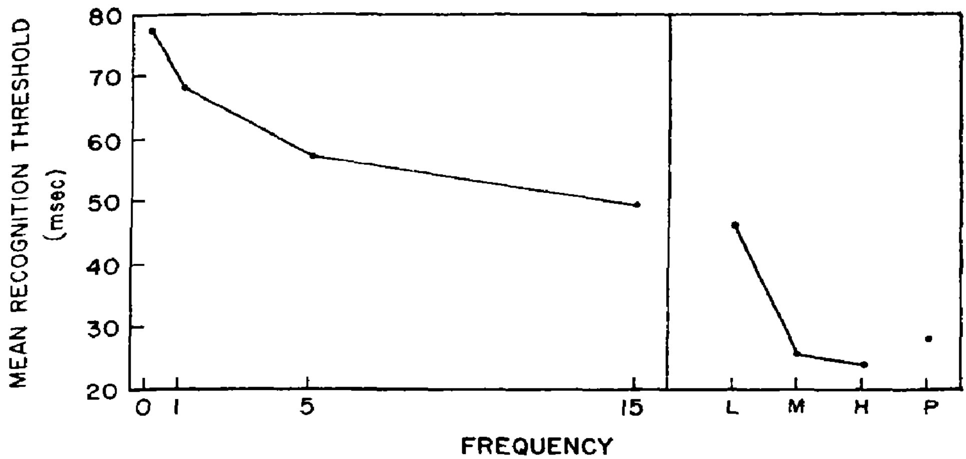


Figure 1. Mean recognition thresholds as a function of frequency and type of word. Numbers refer to artificial words (frequencies of 0, 1, 5, 15), and letters refer to English words (practice, P; low frequency, L; medium frequency, M; high frequency, H)

Table 2. Correlations of recognition thresholds to words of different types and frequencies

	English words				Artificial words			
	P	L	M	H	0	1	5	15
English words								
P	—							
L	81	—						
M	91	.85	—					
H	95	.83	96	—				
Artificial words								
0	.63	.73	.65	.63	—			
1	.65	.76	.71	.65	84	—		
5	70	.78	.77	.75	78	.77	—	
15	.70	.77	.78	.75	74	.80	84	—

Note: Letters for English words refer to practice words, P; words of low frequency, L; medium frequency, M; and high frequency, H.

could be labeled a *familiarity* component, on the assumption that low-frequency English words are more familiar than high-frequency artificial words. It might also be interpreted as a *meaningfulness* dimension. What will be called derived (factor) recognition scores were computed for each subject on each of the two unrotated principal components.

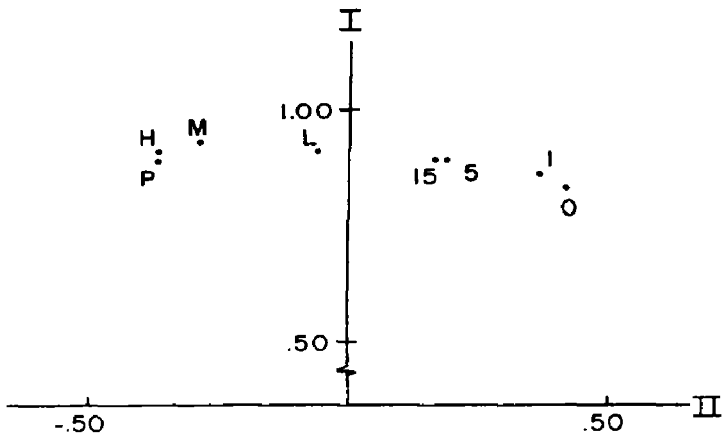


Figure 2. Unrotated principal components of mean recognition thresholds. Numbers refer to artificial words (frequencies of 0, 1, 5, 15), and letters refer to English words (practice, P; low frequency, L; medium frequency, M; and high frequency, H)

Interrelationships of tests and thresholds

The correlations of scores on the five orthogonal test factors with the recognition thresholds to words of different types and frequencies are presented in Table 3. Correlations were also obtained between the derived test and recognition scores. The results are displayed in Table 4. Factor 3, speed of closure, has a significant negative correlation with component I. Factors 2 and 5 have significant negative correlations with component II. A multiple-regression analysis was performed using the derived test scores as predictors and the derived threshold scores in turn as criteria. For component I, no improvement in predictability was found by adding the four other factors to factor 3. Thus, speed of closure was the only predictor influencing overall thresholds. For component II, the multiple R of .38 was significant at the .01 level, with factors 2 and 5 both having significant regression weights, —.22 and —.30 respectively. Thus, a linear combination of memory span and perceptual speed predicts component II. A canonical correlation analysis supported the above results.

DISCUSSION

The hypothesized (test) factor structure is clearly evident despite the ceiling effects on some tests and the low reliabilities of some others. A five-factor solution was obtained, and the Varimax rotation corresponded very well to the hypothesized structure.

Table 3. Correlations of derived test scores and recognition thresholds to English and artificial words at different frequencies

	English words				Artificial words			
	P	L	M	H	0	1	5	15
Factor 1, Word fluency	−.11	−.10	−.08	−.08	−.09	−.05	−.12	−.08
Factor 2, Memory span	−.02	−.07	−.06	−.04	−.23**	−.13	−.20**	−.14*
Factor 3, Speed of closure	−.14*	−.23**	−.16*	−.16*	−.21**	−.18*	−.21**	−.18*
Factor 4, Intellectual performance	.04	−.04	.00	.04	.00	−.01	−.06	−.11
Factor 5, Perceptual speed	.08	−.02	.11	.09	−.15*	−.15*	−.03	−.02

Note: Letters for English words refer to practice words, P; words of low frequency, L; medium frequency, M; and high frequency, H.
* $p < .05$. ** $p < .01$.

Table 4. Correlations of derived test and recognition scores

	Factor				
	1 Word fluency	2 Memory span	3 Speed of closure	4 Intellectual performance	5 Perceptual speed
Component I	− 10	−.12	−.21*	− 02	−.01
Component II	−.01	−.22**	−.07	−.09	−.30**

* $p < .05$. ** $p < .01$.

The inverse relation between word frequency and recognition thresholds for both the English and the artificial words confirms many previous studies. A factor analysis of word-recognition thresholds yielded a two-dimensional solution.

Components I and II were interpreted as an average or overall threshold component and a familiarity component. One of the most apparent features of threshold measurements is their variability across subjects. Speed of closure is one dimension of individual differences that relates to the level or elevation of thresholds. It significantly influences compo-

nent I and is involved in the perception of both English and artificial words at all levels of frequency. Now, the items on the tests of speed of closure were either lowercase words or pictures (each stimulus, then, with its own unique silhouette); the items in the tachistoscopic recognition tasks were uppercase words (each word, then, in the same rectangular shape as the others, so that shape could not serve as a cue). Thus, the perceptual process described by speed of closure and influencing overall thresholds is independent of the shape of the words. Since a figural boundary is generally perceived before figural details (Forgus, 1966), it is evident that this particular process involves mechanisms that do not operate immediately upon receipt of stimulus information by the word-recognition system.

Clearly, other individual differences must enter in to completely account for differences in recognition thresholds. For example, Spielberger and Denny estimate that the correlation of visual acuity and thresholds for high-frequency English words is −.50.

Component II differentiates English from artificial words. The ordered clusters it shows reflect three levels of difficulty in processing a word. On one extreme, high- and medium-frequency English words (including practice words) are familiar; they are units for both perceptual and response systems. On the other extreme, low-frequency artificial words must be perceived letter by letter and built up over a series of exposures: they are not units for either the perceptual or response system. In between, the low-frequency English words and high-frequency artificial words are probably integrated response units; but they are probably processed as less than words. This account is supported by the degree of involvement of perceptual speed and memory span in the data on thresholds. Recall that these two (test) factors are significantly related to (threshold) component II. Memory span is negatively correlated with performance on artificial words but does not relate to performance on English words.[2] Perceptual speed shows significant negative correlations with the low-frequency artificial words and slight positive correlations with the higher-frequency English words. Memory span (the ability to remember strings of unrelated items) and perceptual speed (the ability to compare visual arrays or find elements and figures) are precisely those skills that are required to process unfamiliar material but are *not* necessary to deal with familiar units.

Intellectual performance, a factor on which verbal ability has a high loading, has zero correlations with recognition thresholds of both English and artificial words at all levels of frequency. This is an anomalous result in light of the finding of Spielberger and Denny (1963) and of Richards

and Platnick (1974a): a negative correlation between intellectual ability and recognition thresholds for low-frequency English words was expected.

In the present experiment, the recognition thresholds of the low-frequency English words were slightly lower than the recognition thresholds of the high-frequency artificial words, whereas in Richards and Platnick's (1974a) study, the recognition thresholds of the low-frequency English words were slightly higher than the recognition thresholds of the medium-frequency artificial words. Apparently, the low-frequency words used in the present study were not of low enough frequency in the reading matter of the present pool of students.

Another point deserves consideration. In Richards and Platnick's (1974a) experiment, there were two groups of subjects, half of them with verbal SAT scores of 500 or below and half with scores of 700 or above, whereas in the present experiment, only 9% of the subjects had verbal SAT scores of 500 or below and only 6% had scores of 700 or above. Perhaps a more heterogeneous group of subjects is necessary to detect the correlation between verbal SAT scores and recognition thresholds of low-frequency English words. This was achieved in Richards and Platnick's (1974a) experiment, but not in the present study. There may be a low percentage of subjects who score above or below the critical value. Perhaps there is a minimum verbal SAT score above which subjects have the low-frequency English words in their response repertoire and below which they do not.

Notes

This article is based on a doctoral dissertation by D. M. Platnick, written under the direction of L. G. Richards. The authors thank members of the examining committee (E. A. Lovelace, Herb Richards, Art Schulman, and Wayne Shebilski) for their participation; L. Starling Reid, E. A. Lovelace, Meredith Richards, and David Doggett for their comments on the manuscript; and the Educational Testing Service for permission to reproduce and use the tests from their kit of tests for cognitive factors. The data reported here were also included in a paper read at the Psychonomic Society Meetings in Denver in November, 1975. Offprints may be obtained from Larry G. Richards, Department of Psychology, Gilmer Hall, University of Virginia, Charlottesville, Virginia 22901. Received for publication February 24, 1976.

1. Tables of summary statistics on the tests and detailed reports of the factor analyses are available on request.

2. Recall that recognition thresholds decrease as performance improves, while the various test scores increase. Hence the negative correlations.

References

Cronbach, L. J. 1957. The two disciplines of scientific psychology. *American Psychologist* 12:671–684.

Forgus, R. H. 1966. *Perception.* New York: McGraw-Hill.

French, J. W., Ekstrom, R. B., and Price, L. A. 1963. *Manual for kit of reference tests for cognitive factors.* Princeton, N.J.: Educational Testing Service.

Guilford, J. P., and Hoepfner, Ralph. 1971. *The analysis of intelligence.* New York: McGraw-Hill.

Goldiamond, I., and Hawkins, W. F. 1958. Vexierversuch: The log relationship between word frequency and recognition obtained in the absence of stimulus words. *Journal of Experimental Psychology* 56:457–463.

Howes, D. H., and Solomon, R. L. 1951. Visual duration threshold as a function of word probability. *Journal of Experimental Psychology* 41:401–410.

Kintsch, Walter. 1970. *Learning, memory, and cognitive processes.* New York: Wiley.

Kucera, H., and Francis, W. N. 1970. *Computational analysis of present-day American English.* Providence, R.I.: Brown University Press.

Neisser, Ulric. 1967. *Cognitive psychology.* New York: Appleton-Century-Crofts.

Osgood, C. E. 1964. A behavioristic analysis of perception and language as cognitive phenomena. In *The cognitive processes: Readings,* ed. R. C. Harper et al. Englewood Cliffs, N.J.: Prentice-Hall. (Paper originally published in 1957.)

Richards, L. G. 1973. On perceptual and memory processes in the word-frequency effect. *American Journal of Psychology* 86:717–728.

Richards, Larry G., and Platnick, Daniel M. 1974a. Word-recognition thresholds as a function of verbal ability in two experimental paradigms. *American Journal of Psychology* 87:65–70.

Richards, L. G., and Platnick, D. 1974b. On the influence of pretraining on recognition thresholds for English words. *American Journal of Psychology* 87:579–592.

Solomon, R. L., and Postman, L. 1952. Frequency of usage as a determinant of recognition threshold for words. *Journal of Experimental Psychology* 43:195–201.

Spielberger, C. D., and Denny, J. P. 1963. Visual recognition thresholds as a function of verbal ability and word frequency. *Journal of Experimental Psychology* 65:597–602.

Thorndike, E. L., and Lorge, I. 1944. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University.

AMERICAN JOURNAL OF PSYCHOLOGY
March 1977, Vol. 90, No. 1, pp. 145–156

Sensory information and subjective contour

Edward M. Brussell, Stephen R. Stober, and Deborah M. Bodinger
Concordia University and York University

The possibility that subjective contours are an artifact of brightness contrast was explored. In one experiment, inducing luminance was found to have different effects on the clarity of subjective contours and the magnitude of brightness contrast. The results of a second experiment indicated that differences of luminance in a stimulus display are necessary for subjective contours to be sustained, whereas chromatic differences are not. It is concluded that subjective contour and brightness contrast are distinct perceptual phenomena but share a dependency on the processing of edge information transmitted through the achromatic channels of the visual system.

It is well established that certain stimuli elicit the perception of contours, or abrupt changes in brightness, in the absence of spatially corresponding abrupt changes in luminance (e.g., Kaniza, 1955; Schumann, 1904). Two stimuli that reliably produce such *subjective contours* are presented in Figure 1.

Theoretical accounts of the phenomenon have tended to be either perceptually or cognitively oriented. For example, Coren (1972) argued that the production of subjective contours necessarily depends on the presence of implicit depth cues in the stimulus. He would note of the stimuli in Figure 1 that the 'incomplete' circles and thin-line triangle signal the presence of a solid triangle, an *apparent* triangle, interposed between the observer and the incomplete figures. Similarly, Gregory (1972) argued that information from each configuration is cognitively organized so as to elicit the perception of a solid triangle overlying and masking portions of the circles and thin-line triangle. In either case, the subjective contours appear in order to provide the necessary spatial separation between the apparent triangle and its immediate background.

A different account has been proposed by Brigner and Gallagher (1974) and can be explained with reference to Figure 1. Most observers report the presence of a central triangle that is lighter than its immediate white background for the stimulus at the left of Figure 1 and darker than its

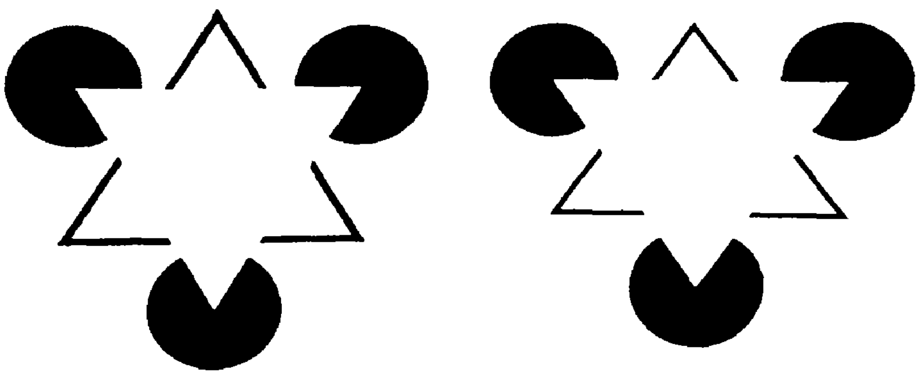


Figure 1. *Two stimuli that elicit subjective contours* (*adapted from* Kaniza, 1955)

immediate black background for the stimulus at the right. The brightness of the apparent triangle relative to that of its immediate background seems to depend on the direction of the physical contrast between the incomplete figures (inducing field) and the remainder of the display (test field). This relation between physical contrast and brightness has been taken to suggest that the appearance of subjective contours may simply reflect the operation of *simultaneous brightness contrast.*

According to Brigner and Gallagher (1974), the stimulus at the left of Figure 1, for example, should elicit a nonuniform brightness contrast, which would result in differences in brightness within the test field. The magnitude of contrast would be strongest within the angular regions of the circular inducers. "By juxtapositioning the areas . . . of comparable apparent brightness, the perception of subjective contour is evoked" (p. 1048). Thus, one should be able to vary the magnitude of the difference in brightness that defines a subjective contour by manipulating those parameters known to influence the magnitude of brightness contrast.

EXPERIMENT I

In order to explore the contrast interpretation of the phenomenon, judgments of brightness were obtained from stimulus regions on both sides of subjective contours as a function of the luminance of the inducing field. In addition, judgments of brightness were also obtained from corresponding regions of a control stimulus that did not elicit subjective contours.

METHOD

Observers

The observers were 20 undergraduates. All had normal or corrected-to-normal visual acuity.

Apparatus and stimuli

Two channels, which we will refer to as I and II, of a Scientific Prototype tachistoscope (model GB) were modified to allow for haploscopic brightness matches to be made. Specifically, baffles were appropriately placed so that the left half of channel I was always viewed by the left eye and that the right half of channel I and the left half of channel II were always viewed by the right eye. In order to prevent the dichoptic images from perceptually fusing, the light entering the right eye was displaced 3.42 deg into the nasal periphery with a rotary prism set at 6 diopters. This arrangement also demanded saccadic eye movements to be made between the dichoptically presented fields, which reduced the probability of the development of afterimages and fading of the images (i.e., Troxler's effect) in either eye.

The control and experimental displays, which are depicted at the upper left and right respectively of Figure 2, were always presented in the left half of channel I (i.e., to the left eye). Each configuration was made from Munsell N 5.0 paper with holes appropriately cut into the surface to create the circular (34.2 min of visual angle in diameter) and linear (31.4 min in length) inducers. The Munsell paper was front illuminated to a luminance of 17.13 cd/m² and had a *gray* appearance. The holes in the paper (i.e., the inducers; see the arrows in Figure 2) were back illuminated and had an achromatic appearance that depended on the ratio of their luminance to the fixed luminance of the gray test field. Between the back illuminators and the Munsell paper were a diffusing surface, to aid in achieving a uniform luminance across the inducers, and Wratten neutral-density filters, to provide control over that inducing luminance. The distance between the centers of the circular inducers subtended 1 deg of visual angle, and the dimensions of the gray test field were 4.3 by 3.3 deg.

The comparison field, which is depicted at the bottom of Figure 2, was always presented to the right eye. It consisted of a circle, 1.17 deg in diameter and of adjustable luminance, against a *gray* background whose luminance was fixed at 17.13 cd/m². In order to have independent control of the luminance of the comparison field and its background, the entire field was created by optically superimposing images from the right half of channel I and the left half of channel II. Specifically, the right half of channel I contained a piece of front-illuminated Munsell N 5.0 paper (4.3 by 3.3 deg). A hole, 1.17 deg in diameter, cut into the center of this surface led into a lighttight enclosure. The left half of channel II contained a piece of black paper with a hole, also 1.17 deg in diameter, cut into its center and back illuminated. When superimposed, the two fields should have produced a comparison field whose brightness could be adjusted higher or lower than that of its gray background. However, because it was extremely difficult to precisely align images from the two channels, a brightness gradient appeared across the comparison field. It was found that placing a thin concentric black ring of 50 min in diameter within the comparison field (see Figure 2) and temporally modulating the comparison field with a 700 msec

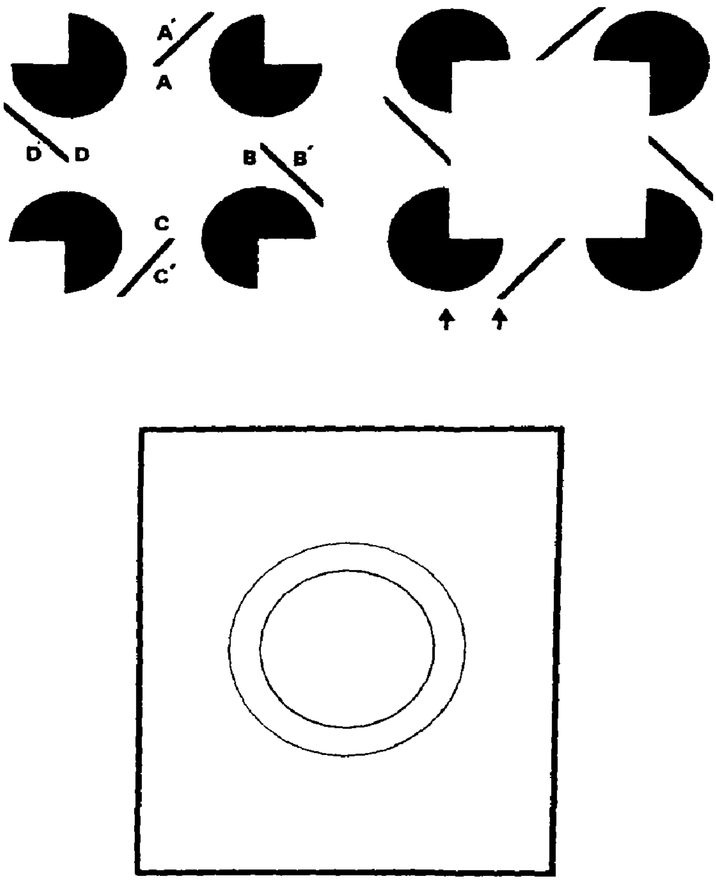


Figure 2. The configurations used in Experiment I. At the upper left and right are the control and experimental displays respectively. The arrows indicate the inducers and the letters the locations on the control display (*inside,* A, B, C, D; *outside,* A′, B′, C′, D′) to which those brightness matches had to refer. At the bottom of the figure is the comparison field and its background. (Note: the achromatic appearance of the three configurations did not correspond to that illustrated and the background of the comparison field is not drawn to scale)

on/50 msec off pulse, resulted in the perception of a uniform distribution of brightness within the inner ring.

Procedure

Before beginning an experimental session, each observer was acquainted with the technique for making brightness matches by having him adjust the luminance of the comparison field until the brightness within its inner ring matched that of a number of achromatic squares. Throughout the experiment, all matches were made using the method of adjustment, and the starting luminance of the comparison field was always randomly determined. The observer was then shown

the experimental and control displays and was carefully instructed about the locations, inside or outside, to which the brightness matches would have to refer. After a series of practice trials, the observer was asked to freely view the display for 5 min in order to stabilize the state of adaptation of his eye (the comparison field was also set to 17.13 cd/m²).

Each of the 24 trials in which the observer participated (6 inducing luminances × 2 judgmental locations, inside or outside, × 2 displays) consisted of a brightness match to one of the four 'inside' or one of the four 'outside' regions. Four blocks of trials, each consisting of one of the four possible combinations of judgmental location and display, were counterbalanced with respect to display. Within each block, an observer made one brightness match at each of six levels of inducing luminance (.9, .6, and .3 log units above and below 17.13 cd/m²), which were presented in a random order. The particular region of the test field (e.g., A, A′, B, B′, etc., at the upper left of Figure 2) to which a given brightness match would have to refer was also randomly determined.

RESULTS

The mean data from the control and the experimental displays are presented at the left and right respectively of Figure 3. In each panel, matched luminance is plotted against log relative inducing luminance, with judgmental location as the parameter.

It might first be noted that none of the mean matched luminances reached a level equal to that of the test field (17.13 cd/m²). This dis-

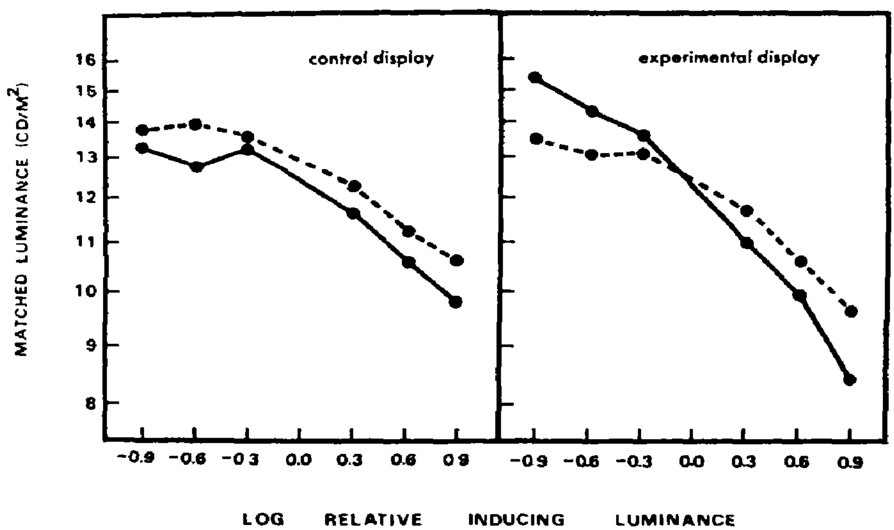


Figure 3. Matched luminance in Experiment I as a function of log relative inducing luminance (at 0, inducing luminance is equal to the luminance of the test field) and judgmental location (*inside*, solid lines; *outside*, dashed lines)

crepancy may have been due to the flickering of the comparison field and/ or the general configurational differences between the test and comparison fields. In any case, prime concern rests not with the absolute level of the matches but with the relationship between the matches as a function of the variables that were manipulated.

A repeated-measures analysis of variance, with three fixed factors, was performed on all of the data.[1] Although the only significant main effect was that of inducing luminance [$F(5, 95) = 36.48, p < .001$], all four interactions were significant [$p < .005$]. These statistical results prompted further exploration into the higher-order main effects and interactions.

Consider first the data from the control display. The effect of inducing luminance was significant at both judgmental locations [inside, $F(1, 95) = 29.34, p < .001$; outside, $F(1, 95) = 29.87, p < .001$], with no significant interaction between inducing luminance and judgmental location [$F(1, 95) = 3.34, p > .05$]. In short, the prominent feature of these data is that they exhibit a distinct effect of brightness contrast, whereby the brightness of the test field varied inversely with log relative inducing luminances greater than zero (see Horeman, 1963). The magnitude of the obtained effect was smaller than that typically found, but this was undoubtedly due to (a) the restricted range of inducing luminances used, (b) the fact that the inducers did not totally surround the test field, and (c) the fact that the comparison field was presented against an illuminated, rather than dark, background.

The effect of inducing luminance at the outside judgmental locations of the experimental display could also be described as a contrast effect. That is, the interaction between the matches at outside locations across displays was not significant [$F(1, 95) = 1.82, p > .05$]. However, the effect of inducing luminance clearly did interact with display for the matches at inside locations [$F(1, 95) = 46.41, p < .001$], indicating that the brightness of the apparent square did not vary in a way that is predictable from the literature on contrast.

The interaction between inducing luminance and judgmental location within the experimental display was significant [$F(1, 19) = 40.98, p < .001$]. This finding lends support to the observers' verbal reports of the presence of a square that was lighter than its immediate gray background when the inducing luminance was lower than that of the test field and darker than its immediate gray background when the inducing luminance was greater than that of the test field. With respect to this interaction, the results of Tukey tests revealed that the curves diverged when the log relative inducing luminance was less than zero, $-.9$ versus $-.3$ [$q(6, 95) =$

4.42, $p < .05$], but were of the same shape when the log relative inducing luminance was greater than zero, .3 versus .9 [$q(6, 95) = 1.96, p > .05$].

One totally unexpected finding was the significant effect of judgmental location within the control display [$F(1, 19) = 7.59, p < .05$]. Since no differences in brightness were reported between these two areas of the control display, this suggests that all of the data may, in part, be reflecting constant error. Although the source of this error is puzzling, it is not problematical. That is, an appeal to constant error cannot explain the *differences* among the data from the control and experimental displays.

DISCUSSION

When the inducing luminance was lower than luminance of the test field, the clarity of the subjective contours varied inversely with inducing luminance. However, under these conditions a contrast interpretation of the phenomenon would predict no changes in brightness within this test field (Horeman, 1963; also see the left panel of Figure 3) and hence no changes in the clarity of the subjective contours. When the log relative inducing luminance was greater than zero, further increases in inducing luminance resulted in greater contrast between the test and inducing fields (see the left and right panels of Figure 3). However, the apparent square and its immediate gray background darkened proportionately, indicating that the apparent-contrast ratio defining the subjective contours remained constant. Under these conditions, the contrast and contours phenomena were dissociated, in that changes in the magnitude of contrast were *not* accompanied by changes in the clarity of the subjective contours.

The data from this experiment imply that subjective contours are not an artifact of brightness contrast.[2] Yet it would probably be a mistake to conclude that the two phenomena are entirely unrelated. It seems likely that they share some processing in common.

There are a number of compelling demonstrations that edge information plays a critical role in the determination of absolute brightness (e.g., Cornsweet, 1970; O'Brien, 1958). In fact, it is this dependency of brightness on edge information that serves as a basis for the prevailing theoretical account of brightness contrast (Ratliff, 1965). The data presented in the right panel of Figure 3 suggest that the determination of different levels of brightness serving to define the subjective contours also depends on the presence of real edge information within the display. Note that the two curves cross at a point that is negligibly different from a log rela-

tive inducing luminance of zero. That is, in the absence of differences in luminance between the test and inducing fields, no subjective contours would have been reported.

To the extent that both subjective contours and brightness contrast depend on the presence of edge information, the production of each phenomenon must depend on the operation of those neurophysiological mechanisms that mediate the transmission of edge information through the visual pathways. In brief, the sensory basis of both phenomena must involve peripheral processing via the operation of lateral inhibition. As manifest in the difference between the data of the left and right panels of Figure 3, the distinction between the two phenomena seems to rest with the manner in which the available edge information is further processed at the central levels of the visual system.

EXPERIMENT II

Depending on whether changes in stimulation are defined by achromatic or chromatic differences, edge information can be transmitted to at least the lateral geniculate nucleus through either of two neural channels. Information about achromatic differences is primarily transmitted within neural pathways whose receptive fields consist of antagonistically organized centers and surrounds with the same spectral sensitivity. Information about chromatic differences is transmitted within neural pathways whose receptive fields consist of antagonistically organized centers and surrounds with different spectral sensitivities (De Valois, 1972).

The data of Experiment I argue the importance of edge information via the achromatic pathways for the production of subjective contours. The question to which Experiment II was addressed was whether information via the chromatic pathways can also serve as a sensory basis for the phenomenon.

METHOD

Observers

The observers were 18 undergraduates. All had normal or corrected-to-normal visual acuity and normal color vision as tested by the Ishihara plates.

Apparatus and stimulus

One channel of a Scientific Prototype tachistoscope (model GB), with front- and back-illuminating capabilities, was used for all presentations. The display was identical to the one depicted at the right of Figure 2. The luminance of the front-illuminated test field was held constant at 17.13 cd/m^2 and had a gray

appearance. Between the back illuminators and the Munsell paper were a diffusing surface and green filter (Edmund Scientific 871). The luminance of the green inducing field could be varied by the observer.[8]

Procedure

For each of the four subjective contours, the observers were instructed to manipulate the luminance of the inducers until one of three situations prevailed: the subjective contour disappeared (*equal* condition), the apparent square was just noticeably lighter than its immediate background (*lighter* condition), and the apparent square was just noticeably darker than its immediate background (*darker* condition). The starting luminance of the inducers was always randomly determined.

Each observer made 12 settings (4 contours × 3 conditions) of the binocularly viewed display. The three conditions were presented consecutively, but randomized for a given contour. The order in which the contours were presented was randomized within observers. All observers participated in a number of practice trials and were light-adapted to the display at a luminance of 17.13 cd/m^2.

RESULTS

The inducing luminance needed for each criterion response per trial was taken to be the mean luminance of the two inducers adjacent to the subjective contour under consideration. The luminance necessary to achieve the criterion response in each condition was represented by the mean of the inducing luminances computed at the four contours for each observer. Thus, the luminance values for the lighter, darker, and equal conditions were directly evaluated, the means being 15.35, 20.14, and 17.68 respectively. In addition, we calculated from the data the inducing luminance at which no subjective contours should appear if only information from the achromatic pathways was important. Specifically, this 'inferred' luminance was taken to be the geometric mean of the settings under the lighter and darker conditions for each observer, the value under this *inferred* condition being 17.58.

The most important finding was that the inducing luminances under neither the equal nor inferred conditions significantly differed from the test field's luminance of 17.13 cd/m^2 [$t(17) = 1.38, p > .05$, and $t(17) = 1.25, p > .05$, respectively]. This strongly implies that chromatic differences between the test and inducing fields were not sufficient to sustain the subjective contours. The phenomenon apparently depends on information transmitted through the achromatic channels.

Data from all four conditions were subjected to a one-way repeated-measures analysis of variance, which revealed a highly significant effect of condition [$F(3, 51) = 75.88, p < .001$]. Post hoc Tukey tests further

indicated that the settings under the lighter condition were significantly different from those under both the equal and inferred conditions [$q(4, 51) = -10.17, p < .005$, and $q(4, 51) = -9.77, p < .005$, respectively], as were those under the darker condition [$q(4, 51) = 11.14, p < .005$, and $q(4, 51) = 11.54, p < .005$, respectively]. These findings further confirm the dependency of the detection and direction of 'subjective brightness differences' on the presence and direction of differences in luminance in the display. The luminances under the equal and inferred conditions did not significantly differ [$q(4, 51) = .39, p > .05$].

DISCUSSION

As a check on the conclusion that information via the chromatic pathways is not utilized in the production of subjective contours, the following informal experiment was conducted. Eight observers viewed the display with three sets of chromatic inducers (green, red, and blue) whose luminance was set at least a half log unit above and below the luminance of the test field. In addition, the entire field was temporally modulated at 3 Hz in order to prevent the development of afterimages. Although a hue complementary to that of the inducers was occasionally seen in the test field, no observer reported any differences in hue or saturation between the apparent square and its immediate background. However, differences in brightness were always reported, differences consistent in direction with what one would expect from the right panel of Figure 3 and the main findings of Experiment II.

GENERAL DISCUSSION

The major conclusion to be drawn from these two experiments is that subjective contours are not a perceptual consequence of brightness contrast. Rather, they share with the contrast phenomenon a dependency on the transmission of edge information via the achromatic channels of the visual system. What is the nature of the critical central process that operates on the transmitted edge information?

Both Coren (1972) and Gregory (1972) imply that the critical central process must be one concerned with organizing the available edge information into spatially distinct regions of figure and ground. According to this view, spatial organization of the sensory information precedes, and plays a role in, the determination of levels of absolute brightness within the test field.

The data in the right panel of Figure 3 point to an interaction between

the manner in which spatial organization affects brightness and the nature of the sensory information being spatially organized. That is, the magnitude of the difference in brightness defining the subjective contours was, under some conditions, influenced by the magnitude of the physical contrast within the experimental display. A theoretical account of this finding awaits a precise specification of the role played by spatial organization in the determination of levels of absolute brightness.

Notes

This research was supported by the National Research Council of Canada (Grant A9937). It is the third author who is at York University. Requests for offprints should be addressed to E. Brussell, Department of Psychology, Sir George Williams Faculty of Arts, Concordia University, 1455 de Maisonneuve Boulevard W., Montreal, H3G 1M8, Quebec, Canada. Received for publication November 18, 1975.

1. For all analyses, the seven partitioned error terms, rather than a single residual error term, were used.

2. Coren and Theodor (1975) have recently also argued against a contrast interpretation of subjective contours.

3. All luminances were measured with a Spectra Brightness Spotmeter, whose spectral sensitivity function most closely matched the C.I.E. photopic function at middle wavelengths. Since it was necessary to accurately measure the luminance of chromatic regions of the display, only green filters were used in the experiment.

References

Brigner, W. L., and Gallagher, M. B. 1974. Subjective contour: Apparent depth or simultaneous brightness contrast? *Perceptual and Motor Skills* 38:1047–1053.

Coren, S. 1972. Subjective contours and apparent depth. *Psychological Review* 79:359–367.

Coren, S., and Theodor, L. H. 1975. Subjective contour: The inadequacy of brightness contrast as an explanation. *Bulletin of the Psychonomic Society* 6:87–89.

Cornsweet, T. N. 1970. *Visual perception.* New York: Academic Press.

De Valois, R. 1972. Processing of intensity and wavelength information by the visual system. *Investigative Ophthalmology* 11:417–426.

Gregory, R. L. 1972. Cognitive contours. *Nature* 238:51–52.

Horeman, H. W. 1963. Inductive brightness depression as influenced by configurational conditions. *Vision Research* 3:121–130.

Kaniza, G. 1955. Marzini quasi-percettivi in campi con stimolazione omogenea. *Rivista di Psicologia* 49:7–30.

O'Brien, V. 1958. Contour perception, illusion and reality. *Journal of the Optical Society of America* 48:112–119.

Ratliff, F. 1965. *The Mach bands: Quantitative studies on neural networks in the retina*. New York: Holden-Day.

Schumann, F. 1904. Einige Beobachtungen über die Zusammenfassung von Gesichtseindrücken zu Einheiten. *Psychologische Studien* 1:1–32.

AMERICAN JOURNAL OF PSYCHOLOGY
March 1977, Vol. 90, No. 1, pp. 157–164

Output interference and intralist repetition in free recall

John H. Mueller and Sam C. Brown
University of Missouri at Columbia

The effect of the repeated presentation of some items in a free-recall list was examined as a function of instructions to recall repeated or unrepeated items first on tests. Instructions to delay the recall of unrepeated items further suppressed their recall, presumably because of increased output interference from the repeated items. However, instructions to recall the unrepeated items first did not enhance their recall as would be expected if output interference was reduced by delaying the recall of repeated items. The role of output interference in explaining the suppressed recall of unrepeated items in the A + 2B paradigm seems limited.

Tulving and Hastie (1972) observed that the recall of once-presented (A) items was suppressed by the presence of repeated (B) items in a list for free recall. This effect was demonstrated by comparing the probability of recall of A items presented with repeated B items (A + 2B) to the probability of their recall in control conditions where new (C) items replaced the repetition of the B items (A+B+C) or where the repetition of the B items was just deleted (A+B). Tulving and Hastie concluded that the 'weak' A items were inhibited in recall by the 'strong' B items.

The present experiments were designed to examine the role of output interference in producing this effect. Previous research (e.g., Tulving and Arbuckle, 1966) has shown that the recall of an item may interfere with the recall of subsequent material. One possible explanation of the suppressed recall of the unrepeated A items in the A+2B situation is that the output of the repeated items serves as a source of interference. This possibility has further appeal in view of the fact that repeated items tend to be recalled before unrepeated items in output (e.g., Tulving and Hastie, 1972, p. 299).

Since Tulving and Hastie's (1972) procedure actually required subjects in the A+2B situation to write the repeated items twice during

recall, it seems especially likely that output interference could be involved in the explanation of the suppressed recall of unrepeated items. Hastie (1975) varied the nature of the instructions for recall, including having some subjects write down all items only once, whether once- or twice-presented. The results of these variations in the procedure for recall led Hastie to conclude that output interference did not provide a complete explanation of the inhibition effect.

There is, however, another variation in the procedure for recall that should affect the magnitude of output interference. This variation involves instructing subjects in the *order* of recalling repeated and unrepeated items during the test. An important factor in an explanation by way of output interference is that the interfering material (the repeated, B, items) be recalled earlier than the target material (the unrepeated, A, items), which is what subjects in the A+2B situation usually do (Tulving and Hastie, 1972; Fritzen, 1975; Hastie, 1975). If the order of recall is critical, it would be expected that subjects instructed to *delay* the recall of the twice-presented items until after they have recalled the once-presented items would show less suppression in the probability of recalling the unrepeated items, at least to the extent that output interference is involved. This was the basic rationale for the two experiments reported below. The primary difference between them was the point at which subjects were instructed about the required order of recall for repeated items. In Experiment I, the instructions about the order of recall were given before list presentation, but in Experiment II, after list presentation. Differential attention to repeated and unrepeated items during study (storage) would be less likely with the latter procedure.

METHOD

EXPERIMENT I

Design and subjects

There were 22 subjects in each of five groups. There were two groups with no repeated items in the list (groups $A+B$ and $A+B+C$) and three groups with a list which had half of the items repeated during presentation. The latter three groups differed only in terms of their instructions about the order of recall for the repeated items: to recall repeated items first (group $A+2B'$), to recall unrepeated items first (group $A'+2B$), or with no special instructions about order of recall (group $A+2B$).

Lists

The lists were constructed from 54 high-frequency concrete nouns chosen from Paivio, Yuille, and Madigan's (1968) norms. For discussion, these may be divided into subsets of 18 items each, constituting the A, B, and C items. Each

2-item A+B list was constructed by selecting 6 items from the A and B sub-:ts, with no items repeated during the presentation order. Each of the A+2B sts was then formed by repeating every B item at some point in the input :quence, resulting in an 18-item list. Each of the A+B+C lists was constructed y replacing the repeated B item with a new item from the C subset, resulting ı a sequence of 18 items without any repetitions. There were three versions of ach list, and these were used variously as first, second, or third tasks.

'rocedure

Each subject had three single-trial lists of the same type and under the same ıstructions. Items were presented visually at a 2-sec rate. A written test imme-iately followed each list presentation and allowed 3 sec per item shown, either 6 sec (group A+B) or 54 sec (groups A+B+C, A+2B, A+2B′, A′+2B). ll subjects given repeated items (groups A+2B, A′+2B, and A+2B′) were ɔld at the outset that some items would be presented twice and that they hould write the repeated items twice on the tests. The two groups with special ıstructions (groups A′+2B and A+2B′) were further told, at the very begin-ing, that they should write the repeated items either first or last during recall.

The third test was followed by a series of digit-span tasks, requiring about 4 ıin. All subjects were then required to write down as many items as they could emember from all three lists. In this case, subjects with repeated items did ıot have to write those items twice but were asked to identify them by a check-ıark. This final test of free recall was included to examine the persistence of he suppressed recall of unrepeated items.

:XPERIMENT II

The only procedural difference was that the specially instructed subjects did ıot receive their instructions about output order in recall until *after* the first ist had been presented. These instructions were then given, and they were re-›eated after each subsequent list had been presented. Interpolating the instruc-ions after input reduces the possibility that subjects might attend differentially o the repeated items at *storage*, so that any instructional effects would more learly be attributable to reductions in output interference at the time of *re-rieval*. The other groups also received some additional filler instructions before ach test, the content being a paraphrase of the initial general instructions. In ach case, the delay was about 30 sec. Otherwise, the procedure was the same as n Experiment I. There were 18 subjects per group.

ЯESULTS

:XPERIMENT I

[mmediate recall

Table 1 summarizes the probability of recalling unrepeated (A) and epeated (B) items per group, pooled over all three tasks. For unrepeated tems, the main effect of groups was significant [$F(4, 105) = 8.28$, $MS_e =$ 071], and the main effect of task was significant [$F(2, 210) = 6.29$, MS_e

Table 1. Average probability of recall of unrepeated and repeated items in Experiment I, pooled over all three tasks

	Groups				
	A + B	A + B + C	A + 2B	A′ + 2B	A + 2B′
Immediate recall					
A items	.64	.52	.47	.46	.38
B items	.71	.58	.76	.73	.80
Final recall					
A items	.43	.37	.37	.32	.30
B items	.54	.47	.53	.51	.51

= .030], but there was no interaction of groups by task [$F < 1$]. Relative to performance with no special instructions (group A+2B), instructions to delay the recall of unrepeated items clearly further hindered their recall (group A+2B′), while instructions to recall unrepeated items first seemed to have little effect (group A′+B2). The performance of group A+2B′ was significantly worse than that of both groups A+B and A+B+C [$ps < .05$]; however, groups A+2B and A′+2B were significantly worse than group A+B [$ps < .05$], but not worse than group A+B+C.

The main effect of groups was also significant for the recall[2] of repeated items [$F(4, 105) = 8.84$, $MS_e = .053$].

The priority given the unrepeated items in output was assessed by computing the relative index of priority, *RIP*, developed by Flores and Brown (1974). The main effect of groups was significant [$F(4, 105) = 33.55$, $MS_e = .267$]. Neither the main effect of task nor the interaction of groups by task was significant [$Fs < 2.05$], so only the overall scores will be considered. For group A+2B, the mean *RIP* score over all tasks was −.26 for unrepeated items, while the instructions to group A′+2B to recall unrepeated items first resulted in a score of .37 and the instructions to group A+2B′ to recall repeated items first led to a score of −.65. This suggests that subjects were able to comply with the instructions on order of recall, and while recalling repeated items first was associated with further suppression of the recall of unrepeated items, there was no corresponding improvement in the recall of unrepeated items when recall of repeated items was delayed. While the former result is in accord with an explanation in terms of output interference, the latter is not.

Final recall

There was no significant effect of groups in the data on the final test of recall for either repeated or unrepeated items [$Fs < 1.95$]. In individual comparisons, however, the performance of group A+B was sig-

nificantly better than that of groups A′+2B and A+2B′ in terms of recalling unrepeated items. The probability of recalling unrepeated items was greater for the third list than for the first two lists [$F(2, 210) = 11.04$, $MS_e = .033$], but this was not the case for repeated items [$F < 1.21$].

EXPERIMENT II

Immediate recall

Table 2 summarizes the data on probability of recall for Experiment II. For unrepeated items, the main effect of groups was significant [$F(4, 85) = 6.37$, $MS_e = .068$], as was the main effect of task [$F(2, 170) = 14.79$, $MS_e = .031$], with no interaction of groups by task [$F < 1$]. As was the case in Experiment I, the instructions to group A+2B′ to delay the recall of unrepeated items further suppressed their probability of recall relative to group A+2B [$p < .05$], but the instructions to group A′+2B to give output priority to unrepeated items did not enhance their recall relative to that of group A+2B. The performance of group A+2B′ was significantly worse than that of both groups A+B and A+B+C [ps $< .05$]. Groups A+2B and A′+2B did not differ significantly from group A+B+C. The performance of group A+2B was significantly worse than that of group A+B [$p < .05$], but the performance of group A′+2B was not. All groups that had been given lists with repeated items recalled those 2B items better than the two control groups [$F(4, 85) = 8.74$, $MS_e = .057$] recalled the unrepeated B items.

In this experiment, the instructions should have their clearest effect on retrieval alone for the first test, since thereafter subjects could anticipate at storage what output order might be required and attend differentially during input. The mean probability of recall on that first test for the once-presented items was .50, .56, and .38, for groups A+2B, A′+2B, and A+2B′ respectively, and .60 and .62 for the control groups A+B and

Table 2. Average probability of recall of unrepeated and repeated items in Experiment II, pooled over all three tasks

	Groups				
	A + B	A + B + C	A + 2B	A′ + 2B	A + 2B′
Immediate recall					
A items	.55	.50	.44	.46	.31
B items	.58	.54	.76	.69	.73
Final recall					
A items	.42	.32	.34	.35	.27
B items	.44	.37	.56	.50	.52

A+B+C respectively. Thus, the pooled performance shown in Table 2 is quite similar to the earliest pattern of performance, and in neither case was there any evidence for the enhancement of the recall of unrepeated items when they were given priority.

The *RIP* scores again indicated that subjects complied with the instructions about output order. For unrepeated items, the mean *RIP* scores pooled over the three tasks were −.31, .55, and −.44 for groups A+2B, A′+2B, and A+2B′ respectively.

Final recall

For unrepeated items, the main effect of groups was significant [$F(4, 85) = 2.60, MS_e = .060$], as these items had the highest probability of recall for group A+B and the lowest for group A+2B′, which originally delayed the recall of these items. Individual comparisons showed the only significant differences to involve group A+B versus group A+B+C, and group A+B versus group A+2B′ [$ps < .05$]. For repeated items, the main effect of groups was also significant [$F(4, 85) = 4.32, MS_e = .068$], as groups A+B, A′+2B, and A+2B′ all generally performed better than the reference groups A+B and A+B+C. Individual comparisons showed group A+B to be significantly different from group A+2B [$p < .05$], but not from groups A′+2B and A+2B′. Group A+B+C was significantly different from groups A+2B, A′+2B, and A+2B′ [$ps < .05$], but not from group A+B.

DISCUSSION

The overall pattern of results in these two experiments indicates that while output interference can be manipulated by varying the order of recall of the repeated and unrepeated items, such interference played a limited role in the suppressed recall of unrepeated items by the experimental subjects with no special instructions (group A+2B). This conclusion seems warranted, since delaying the recall of once-presented items further suppressed their recall (group A+2B′), but recalling once-presented items first did not enhance their recall (group A′+2B), even though such items are not usually recalled first in the absence of special instructions about output order. Thus, the special instructions increased recall suppression but seemed unable to reduce it, suggesting that output interference is not involved when no special instructions are given.

How then can the inhibitory effect be explained? Perhaps the most reasonable explanation is that proposed by Fritzen (1975) and Hastie (1975). These investigators found that the effect occurs only when subjects are explicitly instructed to keep track of the items that are repeated

in the list. When the subjects are not thus required to specifically identify those repeated items, the inhibitory effect is eliminated. On the basis of these results, Fritzen and Hastie concluded that the suppressed recall of once-presented items is due not to repetition per se, but instead to the added demand that the subject discriminate repetitions, thus causing him to pay less attention to the unrepeated items. In this regard, we probably obtained similar results in Experiments I and II here, because even though the instructions about output order followed presentation in Experiment II, the subjects in the three main groups of both experiments knew that some items would be repeated and that they had to make that discrimination during study in order to write the items twice during test.

Whether the (hypothetical) decreased attention to once-presented items renders them unavailable, inaccessible, or both is still uncertain. The only data bearing on this question come from Hastie's (1975) study. In addition to testing for the inhibitory effect in immediate free recall, Hastie used a test of delayed recognition. He found that the inhibitory effect in immediate free recall was eliminated in delayed recognition. Since recognition is assumed to be a sensitive measure of item availability, Hastie concluded that the locus of the inhibitory effect is most likely traceable to the inability of subjects to retrieve (access) the once-presented words from memory.

However, two points should be made. First, Hastie measured inhibition in terms of the difference between his experimental conditions and a *token* control (group A+B+C). Second, he did not include a test of delayed *recall*. The latter seems critical, because if the inhibitory effect is, as hypothesized, due to the difficulty of retrieving (accessing) information that is stored (available) in memory, then a test of delayed recall should be maximally sensitive to the inhibitory effect. But this is not the case, as the present results show. The level of final free recall of once-presented items by the uninstructed subjects of group A+2B, was indistinguishable from that of the token control. Therefore, it would appear either that the inhibitory effect is a phenomenon of relatively short duration and restricted largely to immediate tests or that the token control is not sensitive to detecting the inhibitory effect with delayed tests of retention. At the least, the present results seriously question Hastie's conclusion that the inhibitory effect is due to a failure of retrieval rather than to differences in storage.

That the problem outlined above may be one of sensitivity is suggested by the fact that the inhibitory effect was more apparent in final free recall here when comparisons are made relative to the *type* control (group A+B). While there are some problems associated with both controls,

we believe that the type control is the more appropriate of the two. This is because the token control list contains more different words than the A+2B list, while the type control list does not. Since probability of recall is known to vary inversely with list length, the token control is likely to underestimate the level of performance expected in the A+2B situation in the absence of inhibition. Furthermore, the fact that most of the evidence, including that of the present study, indicates that output interference is not a major factor in the inhibitory effect would also seem to argue that the token control is less appropriate.

It seems possible to draw three conclusions from the present results. First, output interference is probably not the major factor contributing to the inhibitory effect in general. Second, it is not possible to specify the locus of the inhibitory effect from the results of experiments performed to date. Finally, whether the suppressed recall of once-presented items is to be regarded as a temporary or persistent phenomenon may depend on the particular control used to assess inhibition.

Notes

Requests for offprints should be addressed to J. H. Mueller, Psychology Department, University of Missouri, Columbia, Missouri 65201. The authors thank Jay Olson, Tom Overcast, and Paul Rossmeissl for their assistance. Received for publication May 18, 1976.

1. All results described as significant involve $p < .05$ or less.
2. This analysis credited recall of a B item just once for groups A+2B, A′+2B, and A+2B′, even though subjects were required to write these down twice.

References

Flores, L. M., and Brown, S. C. 1974. Comparison of output order in free recall. *Behavior Research Methods and Instrumentation* 6:385–388.

Fritzen, J. 1975. Intralist repetition effects in free recall. *Journal of Experimental Psychology: Human Learning and Memory* 1:756–763.

Hastie, R. 1975. Intralist repetition in free recall: Effects of frequency attribute recall instructions. *Journal of Experimental Psychology: Human Learning and Memory* 1:3–12.

Paivio, A., Yuille, J. C., and Madigan, S. 1968. Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology* 76(1, part 2).

Tulving, E., and Arbuckle, T. Y. 1966. Input and output interference in short-term associative memory. *Journal of Experimental Psychology* 72:145–150.

Tulving, E., and Hastie, R. 1972. Inhibition effects of intralist repetition in free recall. *Journal of Experimental Psychology* 92:297–304.

AMERICAN JOURNAL OF PSYCHOLOGY
March 1977, Vol. 90, No. 1, pp. 165–195

Book reviews

Carl P. Duncan
Northwestern University

Recall and Recognition

Edited by John Brown. London: Wiley, 1976. Pp. 275. $24.50.

This is an impressive book that deserves a widespread readership, for it demonstrates conclusively that recall and recognition are far more than the two differentially sensitive measures of retention or forgetting they have traditionally been thought to be. The differences and similarities of recall and recognition are shown not only to be of general theoretical significance for the understanding of memory but also to contribute importantly to such other substantive research areas as imagery, interference theory, and amnesia.

Editor John Brown of the University of Bristol, a prominent English psychologist perhaps best known as the originator of the Brown-Peterson technique for studying short-term memory, contributes an illuminating first chapter concerned primarily with the theoretical and methodological complexities facing anyone attempting to compare recall and recognition, as procedures or processes. This chapter could profitably be reread after all of the other eight chapters to provide some closure on the current state of the topic and problems.

Brown's assemblage of eleven contributors (all but one of them, Leo Postman, presently located in England or Canada) offer effective antidotes to those provincials who mistakenly view modern cognitive psychology as primarily an American phenomenon. Although quite short (only 242 pages of text plus nearly 600 references) and unfortunately delayed in its publication, the book's contents are both wide-ranging and useful. Chapters 2 and 3 describe some important developments deriving from such controversial topics as encoding specificity (Endel Tulving) and depth of processing (Robert Lockhart, Fergus Craik, and Larry Jacoby). Two other chapters consist primarily of spirited and effective defenses of recently beleaguered dual-process imagery (Allan Paivio) and interference theory (Leo Postman). Useful reviews of the historical and current literature, combined with brief reports of unpublished doctoral dissertations directed by other contributors, constitute two additional chapters. One concerns the roles of study as compared with test trials, coauthored by Aldwyn Cooper and Andrew Monk (under Brown), while the other is about differential effects of word frequency in recognition and recall, by Vernon Gregg (under Craik). The final two short chapters cover memory disorders and presumed physiological mechanisms based on research with human amnesics (Elizabeth Warrington) and hippocampal-lesioned animals (David Gaffan).

This book requires, as does the literature on recall and recognition, that the

reader cope with somewhat inconsistent terminologies and definitions, and not expect to find simple answers on even the basic differences (if any) between recall and recognition. Brown notes that the typically straightforward operational distinction between the procedures of recall and recognition tests need not imply a corresponding difference between recall and recognition as processes. He further contends that recognition tests often involve processes characteristic of recall. Even stronger arguments are presented by others for a continuity of recall and recognition: for example, that they differ only in that 'copy cues' in recognition provide better information from which the initial encoding can be retrieved (Tulving) or reconstructed (Lockhart et al.), or that their processes merely reflect different aspects of the same system (Gregg).

Other authors, however, argue for more basic differences between recall and recognition, differences which Paivio suggests to be localized in retrieval rather than storage, while Postman's distinction is based on different attributes for recall (associative) and recognition (discriminative) during encoding as well as retrieval, and Cooper and Monk describe recall and recognition respectively as retrieval and discrimination tasks. Gaffan focuses explicitly on differences in recognition (but not recall) as aided by 'familiarity discrimination.' Even at the procedural level, both Tulving and Warrington make effective use of intermediate tasks where only some portion of the target item is presented (although Brown characterizes this as recall rather than recognition unless "one or more possibilities are provided for the missing part").

Additional terminological innovations and complexities not surprisingly center around Tulving, who introduces 'ecphory' to refer to "the process by which information stored in a specific memory trace is utilized by the system to produce conscious memory of certain aspects of the original event," which is said to require information both from the memory trace and the 'ecphoric cue.' His alternative to two-stage theory thus is labeled 'episodic ecphory,' although Lockhart et al. suggest substantive changes in Tulving's earlier influential distinction between episodic and semantic memory. They characterize episodic memory as a structureless result of pattern recognition and cognitive encoding operations (it has no built-in time markers and one need only scan it to select relevant items) and semantic memory as the means by which one analyzes and interprets events as they occur in episodic memory. Lockhart et al. further redefine learning (changes in structure of semantic memory produced by encoding operations) as distinguished from memory (encoded features in episodic memory). Postman's warning that "no basic advance is made by recasting old and poorly understood problems in new and timely terms" may well be especially appropriate here.

Another important attribute of this book is its focus on two-stage theory, which is consistently criticized as inadequate and not explicitly advocated by any of the present authors. This two-stage theory (which Tulving alleges to be responsible for his substitution of 'ecphory' for 'retrieval,' because the two-stage theory used the latter term to label its first stage) is instead spoken of in terms like generation/discrimination (Brown), generate/recognize (Gregg), and search/decision (Lockhart et al.). The two-stage theory is most readily identified in terms of its advocates (e.g., Kintsch, Bahrick, Anderson and Bower), and the present criticisms regrettably focus primarily on its earlier and simpler versions. Consequently, a contribution from some advocate(s) of its more recent versions would have made the book better balanced and more current.

Such terminological trivia, however, must not be allowed to obscure the major message(s) of this book, which this reviewer finds to be of paramount significance and relevance to a variety of problem areas. At the risk of misrepresenting the views of some author(s), it appears that at least four quite generally agreed upon conclusions can legitimately be drawn from the book. First, recall and recognition both can represent complex multiprocess phenomena influenced by numerous other variables, with their similarities and differences heavily dependent on specific experimental conditions, methodologies, and instructions. Second, experimental comparisons between recall and recognition, and the effects of other variables on them, contribute importantly not only to the understanding of recall and recognition but also to other topics and issues in current cognitive psychology. Third, more detailed analyses of subprocesses, at the level of features and/or attributes as well as stages or levels of processing and subjects' strategies, will be prerequisites to any complete understanding of either recall or recognition or the relationships between them. And fourth, the contents of this book represent only the beginnings of the extensive research that will be necessary to answer the questions the book raises.

Serious readers of this book will surely find a wealth of specific ideas or results especially intriguing to them. This reviewer's two top choices are (a) Warrington's evidence that amnesic deficits in recall and recognition can be virtually eliminated if part of the target item is presented for prompted recall, an effect she attributes to amnesic deficits in ability to reject interfering incorrect alternatives and thereby implicates interference as a key factor in recognition as well as recall, and (b) Brown's suggestion that incorrect words of higher judged memorability may be rejected more readily because of the subject's certainty that if such memorable words had been presented they would surely be remembered.

Despite these and numerous other cues for future research, there are reasons for skepticism that this book will actually generate substantial new research activity. This is because its level of complexity is comparable to those that in the recent past have produced mass abandonment of a problem area — presumably because the area had surpassed the stage where simple definitive experiments were possible and what was needed instead was the kind of painstaking detailed experimental analysis that many experimental psychologists avoid like the plague. Moreover, this book offers relatively little in the way of appropriate models for research (as distinguished from theory) to guide future experimentation, partly because its contents comprise far more previously published than unpublished research. Paivio does present too brief a preliminary report of extensive correlational analyses including imaginal and verbal encoding as well as measures of recall and recognition for both pictures and concrete words under incidental and intentional learning (and also various frequency, structural, and imagery ratings of these words), but this is an approach unlikely to appeal to the typical experimenter in the area of cognition. A more appropriate model could have been provided by Cooper's experiment that manipulated factorially five differing proportions and/or sequences of study and recall trials, grouped and ungrouped lists, and immediate and delayed tests (the latter of unspecified delay). Unfortunately, complex results producing a three-way interaction among these variables evidently produced a retreat to a simpler second experiment. It is not even mentioned that said interaction reflects a greater decrement from immediate to de-

layed tests for the condition with five successive initial study trials on a grouped (but not ungrouped) list than for any other condition (Figure 1, p. 147), an effect which cannot be evaluated in Cooper's second experiment. Gregg's previously unpublished dissertation involved the more common and less efficient strategy of several (eight) relatively simple and partially overlapping experiments, which provide limited evidence of conditions (e.g., mixed lists of words of high and low frequency) under which frequency may not have opposite effects on recognition and recall.

Countering the above pessimism about the book's impact on future research on recall and recognition is the fact that it prompted the present reviewer to plan a within-list or item-recall-recognition variable for his next major experimental undertaking. Hopefully, enough others will be similarily affected that there will be another book on "Recall and Recognition" to provide definitive answers to at least some of the important questions raised by this one.

William F. Battig, *University of Colorado*

Field Theory as Human Science: Contributions of Lewin's Berlin Group

Compiled by Joseph de Rivera. New York: Halsted Press, 1976. Pp. 533. $24.50.

Five major studies by students of Kurt Lewin when he was at the University of Berlin have been translated into English and form the core of this book. The studies are "The Resumption of Interrupted Activities," by Maria Rickers-Ovsiankina; "Mental Satiation," by Anita Karsten; "On Relapses in Relearning," by Georg Schwarz; "The Dynamics of Anger," by Tamara Dembo; and "Success and Failure," by Ferdinand Hoppe. All were originally published in *Psychologische Forschung* in the years 1923–1933. They vary in length from about 40 to about 100 pages (of this book). Each study has been shortened slightly by de Rivera; the complete translations are available from University Microfilms (Ann Arbor, Michigan). Several other studies by students of Lewin (such as the well-known work of Zeigarnik) are mentioned briefly in the book and are also available in English translation from University Microfilms.

The five major studies form about 60% of the book. The remaining 200 pages, more or less, consist largely of de Rivera's explanations, extensions, and reinterpretations of Lewinian field theory. The result is that the book is really two books. The translated studies are useful for students of the history of psychology in at least three areas: experimental, personality, and social psychology. Also valuable for history are summaries of the work of other students of Lewin. But de Rivera's own version of field theory takes up far too many pages and makes the book much larger than it needed to be.

C.P.D.

Evolution, Development, and Children's Learning

By Harold D. Fishbein. Pacific Palisades, Calif.: Goodyear Publishing, 1976. Pp. 332.

Traditionally, the concept of evolution has been used as a tool for better understanding the social and biological functioning of the adult human being. Specifi-

cally, such topics as intergroup aggression or the size of the occipital lobe of the brain have been studied in terms of their evolutionarily adaptive value. In this recent book, Fishbein turns our attention to a new application of the principles of evolution. He attempts to demonstrate that the developing child may best be understood within an evolutionary, or adaptive, context. By this reorientation, the concept of evolution applies not only to the mature organism but to the maturation process itself.

The goal Fishbein sets out for himself is sizeable, involving a reinterpretation of the literature on child development in terms of step-wise evolutionarily significant stages. The crux of Fishbein's work is that neither environment nor heredity alone can serve as an adequate explanation of the child's behavior. Rather, it is the *interaction* between the two that is critical. The genetic system lays the foundation for directing the developing child within certain constraints, while the environmental factors influence how the child's capacities will emerge from this foundation. The interaction of genetic and environmental influences — referred to as the epigenetic system — theoretically parallels the continual confrontation of genes and environment that is characteristic of the evolutionary process. In other words, the developing child is actually a microcosmic evolutionary analogue, with his or her eventual personality the product of a series of environmental experiences acting to modify the genetic base. The particular concept of maturation as a series of genetically determined choice points with environmental modifications is referred to as canalization.

The evidence Fishbein draws on to support his thesis is necessarily selective. He argues that there are five basic realms of behavior critical to the successful adaptation of man to his evolutionarily most recent mode of existence as a hunter-gatherer: motor skills (to ensure the stability of the social order within the group), language (to communicate with other individuals in the group), attention and memory (to enable better knowledge about the changing features of the environment), spatial abilities (to mentally represent, or schematize, the various locations of food and shelter in the environment), and moral obligations (to ensure the stability of the social order within the group).

Though the evidence is selective, the theoretical superstructure Fishbein develops is highly complex. In fact, it becomes difficult to provide enough empirical support to keep that structure from collapsing. Each of the five behavioral realms is adequately justified as logically useful for appropriate adaptation. However, the specific illustrations given in each chapter are not consistently related back to their evolutionary significance; rather, they are given as examples of behaviors that systematically change in a consistent maturational sequence.

In his discussion of spatial abilities, for example, Fishbein cites four crucial cognitive-perceptual proficiencies: object permanence, perception of single objects, concept of order and straight-line construction, and projective and Euclidian relationships. The author illustrates that within each proficiency, qualitatively different developmental levels may be observed; but he establishes no direct relationship to evolutionary principles. Again, the chapter on language acquisition clearly points out that there is a very systematic and orderly series of stages to describe the growth of vocabulary, the understanding of grammar, and the use of language to regulate (inhibit or initiate) behavior in oneself and others; but the specific adaptive ramifications of these behaviors or the transitions between them are not discussed. Without such ties to an evolutionary background,

Fishbein's orientation cannot be easily differentiated from a strictly genetic one (i.e., Piaget's), which would also predict a series of invariant developmental stages. Since his theory (a variation of the genetic theme) owes its uniqueness to the evolutionary context, Fishbein's facility at explanation should be directed toward the adaptive character of the stages rather than the stages themselves.

Perhaps Fishbein's theory is, in fact, too broad and complex to be encompassed comfortably within a single work. Certainly, the evolutionary framework he proposes is logically and intuitively appealing, yet the manner in which theory and data are juxtaposed is confusing and disappointing. The theory goes in one direction (to argue that development and evolution are intimately related), while the empirical summaries go in another (to demonstrate that there are predetermined stages of performance which occur in a fixed sequence). Their divergence places a heavy burden on the reader to reestablish the relationship. This task should rightfully rest with the author, not the reader.

A more viable approach might be to address a single topic, completely lay out all the relevant data, and demonstrate how each item relates to the evolutionary perspective. This could be done more as a sample than as an all-inclusive summary. Fishbein's section on moral development, in fact, makes use of just such an approach. Each of the topics discussed (obedience and resistance, identification and conscience, empathy, reciprocity, and cognitive/language factors in moral obligation) has an adaptive significance that is clearly pointed out.

An additional problem with the text is the unnecessary inclusion of a great deal of background material on the basic principles of evolution. In order to properly set the stage for his examination of the developmental literature, Fishbein prepares the reader by giving a detailed account of the theories of evolution, as well as the evidence gathered from four main sources: paleoanthropological records, changes in brain structure, social adaptations of species similar to man, and social adaptations of modern hunter-gatherer peoples. The presentation would be smoother had this entire section been deleted and the reader referred to another source for such information. The introductory chapter is excellently structured to arouse one's intellectual curiosity about the specific subtleties of Fishbein's theory. However, by the end of the short course on evolution (six chapters, or about half the remaining book), one's enthusiasm can wane considerably. The reader who is primarily interested in the psychological implications of the theory might skip chapters 2 through 7 and go directly from the introduction to the section on ontogeny. It would be quite practicable to pick up at this point because of the frequent and lucid summaries at the beginning and end of each chapter and at the end of the subsections within chapters. These provide not only helpful reviews of the material but comfortable transitions from one topic to the next.

On the whole, I found Fishbein's book intellectually tantalizing but theoretically confusing. At this early stage of analysis, it would have been valuable to clarify the theoretical issues in depth, leaving the empirical support for a later time. For instance, I was never sure of the exact significance of the transitional stages in the child's development. Does each stage have a survival value for the organism at that point? Is it simply a preparation for a later adaptive behavior? Or is its value not presently adaptive but simply a vestigial holdover of earlier ancestors' adaptive behaviors? Clearly, the answers will distinctly color one's interpretation of the behavior. Fishbein mentions early in the text that these

different possibilities exist (see p. 155), but he never makes an adequately definitive statement on any single position.

In all fairness, Fishbein is aware of the limitations of his endeavor and continually reiterates the blanket assumptions that must be made at this preliminary level of theoretical construction. "Given the present status of canalization studies and ethological studies of human behavior," he says, "some of the discussions in this book will be speculative. Where constancies or invariances in human development are observed, it will be assumed that canalization and induction processes were operating, and that as a consequence the characteristics must have had adaptive value at some point in human evolution" (p. 9).

Despite my criticisms, I firmly believe that this work will have considerable impact on the thinking in this area. It is closely related to the newest developments in the area of animal research on biological constraints on learning and is more remotely related to the current revolt against associationism in the research on human learning. The timeliness, as well as the intuitive appeal, of Fishbein's theorizations are bound to stimulate new interpretations of older phenomena. Sheldon H. White (in the foreword) provides a most eloquent and fitting comment on the work: "It is not enough to discuss children's learning only from the vantage point of the learning laboratory. Science has by now sought out the nature and conditions of human adaptation in many laboratories and field sites." Not only does it take "effort and courage to bring it all together, but this kind of approach and only this kind of approach is in the true spirit of Science. Science cannot be a matter of stockpiling research findings and it cannot be done in the spirit of safety, security, and stolidity. The essence of scientific work is the struggle to find order in nature, to find the pattern of the phenomena. No writing on children's learning begins to compare with this volume in that regard."

Alan S. Brown, *Southern Methodist University*

Artificial Intelligence

By Earl B. Hunt. New York: Academic Press, 1975. Pp. 468. $29.00.

The field of artificial intelligence is the subpart of computer science concerned with understanding the nature of intelligent action and creating computers that can behave more intelligently. Like computer science, it is a mixture of technology and science — of striving to build devices and striving to understand the nature of what has been wrought. It is of fundamental importance to psychologists, for artificial intelligence not only purports to be relevant (as in its stated concern with intelligence) but has in the last decade become a 'foundation science' for cognitive psychology, being the source (and providing continuing technical development) of many important theoretical notions such as the concept of symbolic systems, programmed behavior, plans, semantic nets, heuristic search, goals, means/ends analysis, and control structures.

Earl Hunt provides a competent exposition of the total field. He is a cognitive psychologist (chairman of the psychology department at the University of Washington) who has also worked extensively in artificial intelligence, mostly in concept learning but also in problem solving, pattern recognition, and memory.

Hunt has two viewpoints that structure the book. The first is that artificial intelligence is a congeries of efforts that has no unified theory or even a unifying theoretical program. Just as intelligence is that which intelligence tests measure,

artificial intelligence is that which its scientists happen to do. The second is that the field has a tendency to be too concerned with superficialities such as the latest showy performing program. What is needed is a book devoted to basics — to whatever mathematical and logical structures have been created so far. Thus, the book turns out to be an exposition rather than a textbook: there are no problems given, nor is there any flavor of 'here is the sort of analysis you should learn to do.' It covers the whole of the field in a series of topics, each of which is treated pretty much on its own terms. There is a feeling (to me) of an absence of direction for the reader — of why the book takes up topic X to depth D, and why it presents proofs here but not there. These lacks, though they detract from the book, do follow from its viewpoints; Hunt might counter that they are the fault of the field and not of his choice in designing the book. Furthermore, they do not hinder the book's main contribution, which is to make available in one place and between hard covers an introduction to most of the topics of the field, especially an introduction that is not a popularization but a reasonable technical account.

Part I has two chapters that form an introduction. Chapter 1 is an overview. Chapter 2 attempts to relate some basic notions of computability to the problem of creating intelligent machines (Turing machines, the hierarchy of languages and, correspondingly, of recognizers). While well-intended, it is simply irrelevant to the rest of the book.

Part II is on pattern recognition and has six chapters. It covers what has come to be called classical pattern recognition (which overlaps with statistical decision theory and linear discriminant analysis), as well as sequential recognition procedures, induction of grammars, and the problem of extracting features for use in a recognition system. Sociologically, the fields of pattern recognition and artificial intelligence are somewhat distinct, though both work on aspects of recognizing patterns. It is good to have an exposition that covers the total set of topics reasonably uniformly.

Part III is on theorem proving and problem solving. It has four chapters, one devoted to representation for problem-solving programs, two devoted to heuristic search, and one long one devoted to theorem proving in the predicate calculus (resolution-based theorem proving). This chapter on theorem proving is a good example of one that is technically adequate but approaches being 'just one damn thing after another,' because of the viewpoints of the author.

Part IV is on comprehension. It has three chapters, one on perception (both of vision and of speech), one on question answering, and one on the comprehension of natural language. Most of this part is handled in a more discursive fashion. The chapters on perception and on natural language are particularly weak compared to some of the others in the book. (Again, Hunt comments on this himself at the end, asserting that this follows from the relative immaturity of these areas, which therefore have fewer technical kernels that need exposition.) Part IV also has a final, very short concluding chapter for the book.

A good feature of the book — again totally consistent with the viewpoints adopted — is its use of detailed examples and calculations throughout. Working through these is definitely worthwhile and will be the main vehicle by which the reader will come to grips with the substance of the field. I found chapters 10 (on heuristic problem-solving programs) and 14 (on question answering) particularly good in this respect. Given this, it is unfortunate that the book as a

whole is marred by a relatively high frequency of errors and typos and that some of these interact with the technical presentation.

A drawback of the book — yet again consistent with Hunt's viewpoints — is that there is very little discussion of total systems and how well they perform. Such an evaluation is an important aspect of the field, in my estimation. A good example of the difficulty can be seen in chapter 12, on theorem proving, in which one emerges from an array of techniques, many of which have been described as if they make substantial differences in performance, without any notion whatsoever of what levels of performance can be expected from these techniques or have in fact been demonstrated experimentally.

To date, there are only a few books on artificial intelligence. Hunt's is probably to be recommended over the others for someone who wants a substantial introduction to the field, because of its coverage and because the others that are broad enough in approach are also too popularized. A few topics are missing: automatic programming (the creation of programs by other programs starting with roughly the information that a human programmer might have), some work on induction that does not fit so easily into pattern recognition, and some work in applied areas that raises or illustrates important issues.

A final observation. Curiously, Hunt seems genuinely ambivalent about the whole enterprise (artificial intelligence, more than the book). There is a slight but detectable undercurrent of putting down the field, along with occasional mildly disparaging contrasts of 'men versus machines.' All this, of course, while making a substantial effort to discuss the field in terms of its technical contributions.

Allan Newell, *Carnegie-Mellon University*

Index for the "Archives of Psychology"

Greenwich, Conn.: Johnson Associates, 1976. Pp. 45. Paperback, $25.00.

The *Archives of Psychology* is a journal that was published from 1906 to 1945. It was basically a series of monographs, 300 of them in all. The journal was associated with the psychology department of Columbia University; Robert S. Woodworth was the editor throughout. This book indexes the journal in three ways: it provides (1) a list of all 300 monographs in order of publication, including the number of the monograph, the year, the author, and the title, (2) an alphabetized author index, and (3) a subject index.

It is a lesson in the history of psychology to go through the authors and the titles of the *Archives*. Number 2, in 1907, was Sheperd Ivory Franz's famous study of the functions of the cerebral lobes. Edward L. Thorndike first published in the *Archives* the same year (number 3, on empirical studies in measurement theory); he last published there 31 years later (number 231, in 1938, on the psychology of language). Many famous psychologists had their first or one of their first publications in the *Archives*. Here are some of them in chronological order of publication: S. I. Franz, E. L. Thorndike, F. L. Wells, H. L. Hollingworth, Herbert Woodrow, E. K. Strong, Henry Garrett, Elizabeth Hurlock, Edna Heidbreder, David Wechsler, Elmer Culler, Arthur Jersild, Otto Klineberg, Anne Anastasi, Richard Wendt, E. E. Cureton, John G. Peatman, Robert MacLeod, Rensis Likert, Solomon Asch, Ross McFarland, Gregory Razran, Anne Roe,

Floyd Allport, Muzafer Sherif, Saul Sells, Robert I. Watson, Donald M. Johnson, S. Stansfeld Sargent, Nathan Schoenfeld. At least as many more could be listed.

Many studies in the *Archives* are still worth reading. It is useful to have this index.

C.P.D.

Origins of Intelligence: Infancy and Early Childhood

Edited by Michael Lewis. New York: Plenum Publishing, 1976. Pp. 413. $17.50.

The *Origins of Intelligence* examines infant intelligence from multiple perspectives: biological, social, cognitive, and affective. The editor, Michael Lewis, refrains from summarizing or interpreting the many, many ideas presented in the thirteen papers, preferring to permit each reader to integrate and interact with the material. He does, however, include a chapter of his thoughts on the sociopolitical nature of intelligence testing.

The questions that pervade the literature on the intelligence of adults and children are also present for infants. Exactly what is meant by intelligence, and does intelligence reflect the role of nature or nurture? The special characteristics of infancy, however, produce another set of questions. Does the behavior we term 'infant intelligence' relate to the intelligent behavior of older children, and is there a critical period for intellectual development? The *Origins of Intelligence* does not present clear-cut responses to these questions because, of course, none exist. However, this volume does provide evidence that allows the reader to reflect on the questions.

Researchers who study infants, like their counterparts who study older children, do not agree on what fits the label 'intelligence.' Such disagreement is no more evident than when we examine the many tests devised to measure the construct. Two chapters, Brooks and Weinraub's "A History of Infant Intelligence Testing" and Honzik's "Values and Limitations of Infant Tests," point out that while drawing heavily from Gesell, each test constructor emphasized different aspects of intellectual behavior. Griffiths, for example, included twice as many speech items as other testers, and Cattell eliminated any items that tapped 'home training.' When similar content was used, the items were not always placed at the same ages. Brooks and Weinraub report that patting and feeling one's image in the mirror was said to typify a child of six months, seven months, or twelve months, depending on whether the Bayley, Cattell, or Buhler scales were employed.

In "Cross-Cultural Studies of Infant Intelligence," Rebelsky and Daniel emphasize the difficulties of defining intelligence from a cross-cultural perspective. They repeat Labov's concern that 'intelligent qualities' are often those that are prevalent in one's own social group, and they describe the Price-Williams distinction between the *emic* approach (defining intelligence according to criteria of the culture studied) and the *etic* approach (defining it without reference to those criteria) to intelligence testing. As an example of the inherent difficulties of the latter approach, Rebelsky and Daniel discuss the Ugandan tendency to equate intelligence with slowness. How accurate a score would timed IQ tests provide in that culture?

The Piagetian concept of intelligence is presented by Ina Uzgiris in her chap-

ter, "Organization of Sensorimotor Intelligence." Piaget's approach differs from the traditional one in two major respects. First, rather than perceiving the infant's intellectual functioning as a downward extension of childhood functioning, Piaget considers infant intelligence to differ qualitatively. He feels that the infant's intelligence is based on perceptions and movements, develops in discrete stages, and subsequently becomes integrated into higher-level intellectual structures. An infant is judged by his or her position in the developmental sequence, not in comparison with other infants of the same age.

Second, whereas the implicit assumption behind most psychometric tests is that the rate of intellectual growth is constant within infants across ages (which is why researchers expect to find adequate age-to-age correlations), Piaget expects neither a consistent rate of growth throughout infancy nor a consistent rate within a given child. While some discrepancies have been noted, most of the evidence reviewed in this volume supports the Piagetian viewpoint. Not only are age-to-age correlations generally low within infancy and between infancy and childhood (see McCall's excellent summary), but evidence for congruence of stage level across item domains varies. Depending on the characteristics of the sample and the particular test administered, stage designation across item content can be moderately congruent to incongruent. Uzgiris suggests, however, that the lack of congruence may be an artifact of employing subjects within narrow age ranges. In keeping with the Piagetian approach, she suggests grouping infants according to their developmental level on a particular subscale rather than according to age.

Scarr-Salapatek writes on "An Evolutionary Perspective on Infant Intelligence: Species Patterns and Individual Variations." Like Piaget, she feels that the intelligence of the infant is qualitatively different from the intelligence of the child. Unlike Piaget, she does not think that sensorimotor skills are the prerequisites of later symbolic intelligence. According to Scarr-Salapatek, sensorimotor intelligence evolved earlier in human evolutionary history and, unlike more mature forms of intelligence, has resisted cultural influences to alter its structure. Thus, the intelligence of the human infant shows strong commonalities with that of other primate infants. The evidence reviewed suggests that infant intelligence has been highly canalized at the species level and that naturally occurring home environments will not alter the child's ultimate development of sensorimotor skills. This is not to suggest that the environment and learning play no part (for, obviously, children differ in rate of development) but that natural selection may have provided existing genotypes with the preadaptation to respond to learning experiences available in almost all environments.

Rather than responding directly to the question of what intelligence is, Lewis focuses on the nature of the question itself. His chapter is titled "What Do We Mean When We Say Infant Intelligence Scores? A Sociopolitical Question," and in it Lewis reminds us of the historical connection between measures of intelligence and academic achievement. Even for infants, the tests are used mainly to predict achievement or to measure gains after intervention. Yet this use may not be warranted. Noting that the skills learned during an intervention project may not show up in measurable IQ changes, Lewis sardonically cautions that "it may still be necessary to select a carpenter for his carpentry skills rather than his IQ scores." He concludes his examination of the intelligence-testing movement by stating that the IQ test has become the twentieth century's 'stratifica-

tion device,' replacing the class system as a way of determining each child's socioeconomic status and position in society.

Thus, several different conceptions of infant intelligence exist. While most authors avoid committing themselves to a specific definition, they tend to agree that intelligence is not unitary but that the various skills may be tested by observing an infant's interaction with objects and people. Furthermore, the items that appear to tap intelligence in infancy are unrelated to those that do so later in childhood. Therefore, predictability of childhood IQ from infant IQ is understandably low.

Given such low predictability, what is the use of infant testing? One is to assess and understand the infant's current level of intellectual functioning. Hunt's chapter, "Environmental Risk in Fetal and Neonatal Life and Measured Infant Intelligence," describes the relationship between such hazards as drugs during the mother's labor and delivery, prematurity, hypoxia or hyperoxia, and the infant's intellectual growth. Hunt also discusses the difficulties of determining the precise causal factors and effects. Studies of a specific insult and outcome are difficult to control, given possible multiple causation. Although more control may be achieved by identifying high-risk groups, the small sample sizes make specific conclusions difficult.

Another promising use of infant testing is with nonnormal populations. Various writers, such as McCall and Brooks and Weinraub, have noted that prediction of later IQ for infants classified as retarded or neurologically impaired is significantly greater than prediction in normal samples. Thus, among some subgroups the general lack of a relationship between early and late intellectual functioning may not hold.

Then there is the nature/nurture issue. While most of the writers directly or indirectly address this question, Lewis notes that sociopolitical interests are more at the root of the controversy than scientific interests. He repeats Cronbach to the effect that even if the heritability index were .80, environmental differences may have as much as a 25-point impact on IQ. Accordingly, cultural and environmental influences can be highly significant.

Other writers, however, report evidence that heritability is probably considerably less than .80. As with older children, the correlation in IQ between identical infant twins has been found to be substantially higher (.81–.85) than the correlation between fraternal twins (.55–.74). The incidence of retardation among identical twins, however, is twice that among fraternal twins. When Nichols and Broman eliminated retarded children from their sample, they reported equivalent correlations between fraternal and identical twins at eight months of age. These results imply that among normal children of this age, there is no evidence of heritability. While McCall reports that heritability may increase after the first year, the evidence presented in this volume suggests that heritability does not serve a practical limiting function on tested intelligence.

For example, Golden and Birns, in "Social Class and Infant Intelligence," discuss the evidence from studies of enrichment programs such as Heber's Milwaukee Project and the projects of Gordon, Levenstein, Karnes, and others. The results demonstrate that substantial gains (as many as 39 IQ points) can be made when a child is exposed to such intervention. Since most of the programs were costly, they are unlikely to be implemented on a large scale. Golden and Birns suggest, therefore, that researchers attempt to identify the precise factors that caused the improvements. One factor that appears to be effective is ma-

ternal involvement: programs such as those of Karnes and Levenstein, which include the mother in the intervention, have the most long-lasting effects.

One factor that does not appear to be important in determining the success of intervention is the subject's age at entry. Golden and Birns conclude that there seems to be no correspondence between age at the beginning of the intervention and the amount of intellectual gain. These results question the concept of a 'critical period.' Although we are accustomed to thinking that early experience sets the limits of later development, the research does not support this conclusion. In addition to Golden and Birns, Honzik also discusses the resiliency of cognitive development. She reports Kagan and Klein's findings that although Guatamalan infants may spend their days enclosed in dark windowless huts and rarely interact with others, their preadolescent test scores are comparable to American middle-class norms. Results such as these encouraged Rebelsky and Daniel to suggest that we should concentrate less on the liabilities of some early interactions and more on efforts to optimize growth at any stage of life.

The relationship between language and infant intelligence is discussed by McCall, in "Toward an Epigenetic Conception of Mental Development in the First Three Years of Life," and in a more narrow sense, by Blank and Allen, in "Understanding Why: Its Significance in Early Intelligence." McCall notes that students of language no longer regard its acquisition as completely innate, progressing without environmental support. Fels's data suggest that sensorimotor behaviors are "necessary antecedent schema for language to build upon." In the same way that the child produces a set of behaviors that have been associated with an object, the child names objects and their qualities (e.g., *stove, hot*). Thus, language may serve the same purpose as sensorimotor skills, allowing the child to produce consequences in the environment.

The above portrayal of language acquisition is similar to 'concept matching,' by which the child learns words for objects or events already understood on a sensorimotor level. Blank and Allen suggest that such a scheme is useful for language with 'portrayable correlates,' those terms representable in a sensorimotor manner. Terms that are nonportrayable correlates, however, require the child to move beyond sensorimotor functioning and are acquired by 'concept formation.' An example of such a term is *why*. Blank and Allen traced the development, from 18 months on, of its use by a female infant. Their conclusions are similar to those of other investigators who have studied the issue. First, in contrast to most words in the language, *why* is produced before it is understood. Second, in comparison with other 'wh' words, its meaningful use is delayed. Third, children ask *why* questions of adults, but almost never of other children. Through their investigations, Blank and Allen became convinced that with the development of terms such as *why*, the child is propelled beyond sensorimotor thought into a more conceptual style.

It is often said that intelligence is what intelligence tests measure. In the *Origins of Intelligence*, however, the reader gets the idea that even given a broad notion of intelligence, some things the tests measure are not meant to be subsumed under intelligence. For example, Uzgiris and McCall both suggest that attention and alertness may be measured by the infant-intelligence test. In fact, the low age-to-age correlations may be attributable to these ephemeral states.

While some factors measured by the intelligence tests may not be labeled intelligence per se, they may be useful in predicting later intellectual performance. Certain enduring aspects of personality or cognitive style may determine

the child's approach to cognitive tasks regardless of the child's age. Golden and Birns suggest that a simple rating of the pleasure the infant appears to take in doing a task may be more predictive of later IQ test performance than success or failure on the task itself. Likewise, Yarrow and Pederson, in "The Interplay between Cognition and Motivation in Infancy," suggest that motivation may be predictive of intellectual growth. An infant who actively seeks out interaction with the environment may control the reinforcements received. Thus, in a sense, the child controls his or her learning.

Watson's chapter, "Early Learning and Intelligence," suggests another means of predicting later intellectual performance. After he discusses early learning as a determinant of intelligence and intelligence as a determinant of early learning, Watson proposes the use of early learning to predict intelligence. He suggests that the predictability of later ball-throwing capacity may be greater after a training session than before, in that the training session may indicate those children who would profit from experience. So also might an early training session in a cognitive skill allow for differences between infants to emerge. Watson reports that such things as practice effects have been viewed as contaminants to testing rather than as possible aids to prediction.

In "Looking Smart: The Relationship between Affect and Intelligence in Infancy," Haviland suggests that perhaps unconsciously we have been using facial expression as an important cue to the infant's intelligence. Some aspects of facial expression are built into the test procedures (e.g., being 'surprised' at the disappearance of an object, showing 'interest' in one object over another), but others may influence the test score in more subtle ways. Haviland devised an ethogram of the facial expressions of a set of twins. The twins differed dramatically in the frequency of various facial expressions, and the differences influenced adults' descriptions of the children. These results suggest that the infants' affect may determine the social responses of others, perhaps even the IQ tester. The importance of affect is further demonstrated when autistic children are examined. Mothers of autistic children often report that their children lack affect (e.g., 'he never smiled at me'; 'he never noticed when I left the room'). Although much research has been done on the development of the infant's recognition of facial expressions, Haviland's research suggests that the impact of the infant's own expressions needs to be explored.

I highly recommend this volume to researchers in the field of intelligence. Although more careful editing might have eliminated redundant material presented by several authors, by and large the papers contribute an informative summation of current thinking in this area. The authors are especially careful to point out deficiencies in the field of infant intelligence and to suggest new avenues of research.

Sharon D. Herzberger, *Northwestern University*

Theories of Vision from Al-Kindi to Kepler

By David C. Lindberg. Chicago: University of Chicago Press, 1976. Pp. 324. $20.00.

This book covers the history of visual theory from the Greeks, the Islamic scholars, and the Western writers of the Middle Ages and Renaissance, to the development of the theory of the retinal image by Kepler. After an introductory

chapter on the ancients (primarily Plato, Aristotle, Euclid, and Ptolemy), the next chapter goes into some detail on the work of the first major Arab philosopher, Al-Kindi (ninth century). Hunain, Avicenna, and Averroes are covered more briefly. Alhazen, however, is said by the author to be the most important writer on optics between antiquity and the seventeenth century, and a long chapter is devoted to him. Alhazen destroyed the extramission theory, which maintained that something emanates from the eye to the object to produce vision.

The revival of learning in the West, beginning in the eleventh century, included renewed interest in vision. Among others, Robert Grosseteste, Albert the Great, Robert Bacon, Henry of Langenstein, Blasius of Parma, John Buridan, and Nicole Oresme wrote and lectured on the Aristotelian theory of vision and on approaches to vision through the perspectivist (ray geometry) theory. The anatomy and physiology of the eye also received considerable attention. Subsequently, some of the great painters of the Renaissance (Giotto, Brunelleschi, Leonardo da Vinci, and others) developed the theory of linear perspective, which they described in their books and illustrated in their paintings. Lindberg shows that they were familiar with the writings of the major visual theorists.

Finally, in the seventeenth century, Kepler firmly established the theory of the retinal image. It may seem too bad that because his theory was correct, he was forced to wrestle with the fact that the retinal image is inverted and reversed. But in dealing with critics on this issue, Kepler arrived at another truth: optics stops at the retina; after that something else takes over.

Lindberg has written an interesting book that is scholarly without being pedantic. Although he regrets that fact that the footnotes had to be relegated to a separate section (60 pages long) in the back of the book, I think most readers will appreciate that.

C.P.D.

The Use of Definite and Indefinite Reference in Young Children

By Michael P. Maratsos. New York: Cambridge University Press, 1976. Pp. 144. $13.95.

The behavioral basis of this book is ostensibly very simple: the use of *a* and *the* by young children. In fact, at first one wonders how this topic can sustain an entire book. Maratsos quickly demonstrates, however, that measurement of definite (*the*) and indefinite (*a*) linguistic referential abilities in young children is not as easy as it appears. Furthermore, he points out how these abilities interrelate with other important conceptual and methodological issues in developmental psycholinguistics.

Maratsos first introduces the reader to the basic background on the semantics of articles, including how a speaker typically differentiates between situations where specific and nonspecific reference is appropriate (such as conversationally introduced specificity, specificity by entailment, etc.). Early in this first chapter, he confronts the reader with the two central questions addressed by the book. How facile is the young child in handling the distinction between specific and nonspecific class membership? And how important is the child's awareness of the cognitive reference point of the listener (nonegocentricity) in the successful development of his skills of indefinite and definite reference?

The second chapter spells out the need for the present investigation. Previous

research has been done by R. Brown on the use of definite and indefinite reference in a naturalistic setting. The main problem with these investigations is that the majority of the references are to objects in clear sight of both the speaker and listener. It is critical, Maratsos contends, that the entire referential continuum of concreteness to abstractness be examined in order to understand the full scope of the child's capabilities.

To accomplish this goal, he constructs a series of exercises to tap both the concrete (objects physically present to speaker and listener) and abstract (referents introduced hypothetically in a story) ends of the dimension, as well as points between (objects physically available to either the speaker or the listener but not both). In the chapters that follow, Maratsos presents a detailed account of a series of games, stories, and sentence-repetition tasks especially tailored to address these different levels.

His methodology is elaborate and well planned. Each of the three- and four-year-olds (20 at each age level) was visited at home and then run through two carefully counterbalanced experimental sessions. The selection and refinement of the tasks is thoroughly explained, in a very readable manner. A separate chapter is set aside for each type of task, in which he describes the specific stimuli and examines the results derived from that exercise. In the final chapters, he summarizes and integrates the results of the various tasks into a general comment on children's referential abilities.

Several points clearly emerge in his discussion of the performance on the various tasks. First, a child's ability to correctly use specific and nonspecific reference develops early. In fact, soon after children begin correctly differentiating between *a* and *the* (at about age three), they are able to consistently and accurately apply the appropriate rules. "In a large number of experimentally created circumstances, varying in the abstractness and awkwardness of the referential situation, we . . . found that even the knowledge of children who had not been using articles long generalized far beyond the limited circumstances of more natural language" (p. 93). This contrasts sharply with the previous work by Piaget and Bruner, which implied that these abilities do not emerge until a later period.

A second important finding pertains to the degree of egocentric orientation (consideration of only his or her own perspective) or nonegocentric orientation (ability to consider the listener's as well as his or her own perspective) a child has in making references. Apparently, this capacity to take into account the state of the listener's knowledge evolves more slowly than the general competence with definite and indefinite reference. Maratsos adds a word of interpretive caution on this egocentricity issue. It is commonly assumed that a general mediational principle underlies the nonegocentric referential development, which serves as a basis for deriving the specific references made in any given situation However, another possibility exists: the child may actually learn a number of separate subrules to guide referencing behavior without developing the general nonegocentric orientation. As Maratsos points out, it is problematical whether this issue can be resolved with this, or any other, design.

The final point is more methodological than empirical. Repeatedly, throughout the various tasks, Maratsos emphasizes how the effects he obtains are very sensitive to variations in the design of the test. He notes that "the relation of experimental techniques to competencies they measure is clearly a problem. . . We have seen how different tasks may shift the baseline of the definiteness of

children's performance. . . . Such results may serve as a caution against believing that the competencies we study are uniformly defined, homogeneously expressed entities, such that all sorts of operations are equally good for assessing the same competence" (pp. 105–106). The problem is important enough to warrant a small chapter, in which Maratsos explains that how the question is phrased (whether to ask what, who, which, or which one) or how the story is worded has a large influence on the type of response. In fact, there are some tasks on which even the adults (parents) did poorly at using the strictly appropriate referent, with their levels of performance not differing significantly from those of their progeny.

This recurrent sensitivity to method tends to suspend the concerned reader in an interpretive limbo. However, Maratsos comes to the rescue with the continual reiteration of just "how complex some of the 'basic dispositions of the cognitive and perceptual structure' required by ordinary linguistic usage may be" (p. 95). He uses this to justify the investigative paths on which he is embarking as necessary, if somewhat long-term.

It is to Maratsos' credit that he points out methodological stumbling blocks throughout the work. He often introduces clever solutions to these problems too. For instance, when he encountered children who would point at a specific member of a group, thus obviating the need for the appropriate reference, he would request that they not point. If this failed, he would have them keep their hands on top of their head as a new element in the game (and maintain this during successive tasks, if necessary).

Structurally, the book suffers from two problems. First, the organization of the middle chapters on the actual investigation is choppy and difficult to follow in places. The reader must continually shift among method, results, and interpretation. Because each task resulted in slightly different findings (as mentioned earlier), Maratsos has to present each task individually and try to pinpoint the factors that could account for the specific deviations found with that exercise. For instance, the reader may find a story presented and discussed, results analyzed by different age levels to illustrate a particular theoretical point, another story introduced, followed by a methodological consideration. This disjuncture is in marked contrast to the initial and final chapters, which are tightly constructed and smoothly presented.

The other structural problem concerns the excessive amount of repetition and detail. For example, almost every game and story is completely described in the text. A sample of each type would have been appropriate, with all the material presented in an appendix. This point relates directly to a general uneasiness I felt as I read the text. Why a book rather than an article? Being basically a dissertation, it may well have been too long for an article and too short for a book, so that the repetition and detail result from an attempt to extend it to a length more suitable to a book (although the final product is still brief). But my question touches on the broader issue of the problematical nature of the exchange of information in psychology at the present time. With the ever-increasing number of journals in most areas of research, it becomes more difficult to achieve the appropriate emphasis in an article. Although a book may be a means to circumvent this difficulty, it poses additional financial and logistical problems for the reader. An adequate solution to this dilemma will probably not come easy, or soon.

Aside from this issue, Maratsos' investigation has considerable merit. He demonstrates skill and ingenuity in applying experimental techniques to natural language processes. Besides revealing children's early mastery of definite and indefinite reference, his procedures instill a new reverence for the complexity of language acquisition. Even the simplest linguistic phenomena are difficult to delineate or explain with precision. As for any pioneering research, the real genius is in asking the appropriate questions, rather than in finding the answers. Maratsos provides a superior model for such an endeavor.

Alan S. Brown, *Southern Methodist University*

Theories in Contemporary Psychology, 2d ed.

Edited by Melvin H. Marx and Feliz E. Goodson. New York: Macmillan, 1976. Pp. 642. $14.95.

The book contains 45 papers, 39 of them previously published. Of the six new chapters written specifically for the book, four involve one or both of the editors as authors. Most of the authors of papers are psychologists; the second largest group of authors is philosophers. For almost all of the selections, the original date of publication is after 1960.

The papers are divided into three major parts, each with several subsections. Part I, "Theoretical Climate," has 17 papers under subheadings on classical positions, historical perspectives, shifting paradigms, and philosophical perspectives. Part II, "Theory Construction," has 11 papers under sections covering the nature of theory, logical perspectives, and procedural issues. Part III, "Levels of Analysis," has 17 papers under subheadings that provide an overview, discuss the experiential, physiological, behavioral levels, and move toward unification.

It is very convenient to have a book such as this in which diverse theoretical papers are gathered together. But in my opinion, there are too many papers and the diversity is too great. The editors say in the preface that they were forced to reject many articles they would have liked to include. Undoubtedly this was the case, but I still think they tried to cover too many different areas of theoretical psychology and philosophy.

C.P.D.

Motion Sickness

By J. T. Reason and J. J. Brand. New York: Academic Press, 1975. Pp. 310. $24.25.

Certainly, many psychologists will join me in applauding this demonstration that an ancient and ubiquitous 'sickness' can be effectively comprehended using concepts derived from research in experimental psychology and animal behavior. Reason and Brand argue persuasively that motion sickness is a maladaptation resulting from sensory rearrangement and that both the initial disturbance and the subsequent recovery can be understood by applying a reafference-feedback model.

This review is divided into three parts. First, the major points considered in the book are summarized. With this summary I hope to demonstrate that this book is worthwhile reading for psychologists who are interested in general issues

related to learning and perception and that it also contains material which could be used to enliven or vary a lecture. Second, the argument is made that this book can be viewed as part of a developing paradigm which should help move the vestibular senses back into the center of perception research and sensory psychology, where they rightfully belong. Third, a few observations are offered on the special problem of motion sickness in the zero-gravity environment.

Motion Sickness is divided into ten chapters. The historical perspective provided by chapter 1 is an excellent introduction to the remainder of the book. The style is established as scholarly yet easy to read, and the chapter is replete with curious facts and anecdotes. For example, motion sickness was a legal punishment for delinquent youths in Prussia in the mid-nineteenth century. Offenders were placed in a box and rotated rapidly until they yielded up a "disgusting spectacle." Early theories of motion sickness emphasized vascular and gastric changes. The 'abdominal vacuum' theory recommended that before embarking on an ocean journey, people eat pickled onions so that the resulting gas would distend the stomach.

There was a shift away from the 'blood and guts' theories of motion sickness toward the end of the nineteenth century, when the contributions of vestibular receptors and the similarities between motion sickness and Ménière's disease were first recognized. William James, who demonstrated a reduced susceptibility to motion sickness in deaf-mutes (who frequently exhibit labyrinth defects), supported this shift. While nonvestibular theories persist even today, researchers of the twentieth century have provided evidence that an intact vestibular-receptor system is a prerequisite for motion sickness. The chapter ends with a summary of the two major current theories of motion sickness: vestibular overstimulation and sensory conflict.

Chapter 2 briefly examines the extent of the problem of motion sickness. With the decline in commercial passenger travel by ship, the incidence of motion sickness among civilians probably has abated. The military problem, on the other hand, has increased. Ship and airplane personnel frequently are confined during performance of their duties in dimly lit compartments without the ability to refer motion sensations to the outside visual world; this confinement is associated with an increased incidence of motion sickness. That motion sickness is a problem for those who would venture into space is illustrated by the fact that five of the nine Skylab astronauts reported its symptoms.

Behavioral and physiological characteristics of motion sickness are reviewed in chapter 3. The cardinal signs of pallor, cold sweating, nausea, and vomiting are the most reliable and consistent features of motion sickness. Numerous attempts have been made to determine other reliable physiological indicators of motion sickness, but the results of those efforts do not present a consistent picture. Researchers also have had difficulty in developing behavioral techniques for determining motion sickness without influencing its development, for while it is regarded as inhumane to continue experiments with human observers to the point of their vomiting, self-reports force the observers to focus attention on internal states in a manner that is known to potentiate the development of the symptoms of motion sickness.

This chapter also includes an interesting discussion of the relationship between motion sickness and depression. While depression and fatigue among the pas-

sengers have been reported frequently by ships' physicians, these responses can be overcome if individuals are highly motivated. The Skylab astronauts, who exhibited extraordinary motivation, were able to work at a high level of efficiency in spite of their motion sickness.

Observations with labyrinthine-defective human observers and labyrinthectomized animals demonstrate conclusively that the vestibular receptors are necessary for the development of motion sickness. These results are presented in chapter 4, which also includes a discussion of the relative importance of the semicircular canals and the otolith organs (utricle and saccule) for producing sickness. Wendt's studies of motion sickness using a 'wave machine' have been cited to support an otolith-only theory. Because his observers' heads were not restrained, Wendt's data cannot be considered to provide unequivocal support for the otolith theory; rather, these observations can be comprehended more readily within a sensory-conflict framework.

The heart of the book is chapters 5 and 6. The essential nature of the stimuli evoking motion sickness is considered in chapter 5, where the phenomenon of motion sickness is firmly embedded in the theoretical context provided by Stratton (*reversed retinal images*), von Holst and Mittelstaedt (*das Reafferenzprinzip*), and Held (adaptation to sensory rearrangement). Reason and Brand argue that "*all* situations which provoke motion sickness are characterized by a condition of sensory rearrangement in which the motion signals transmitted by the eyes, the vestibular system and the nonvestibular proprioceptors are at variance not only with one another, but also — and this is the crucial factor — with what is expected on the basis of past experience or exposure history" (p. 105).

The chapter is organized on the basis of a six-part classification of the kinds of sensory rearrangement that can provoke motion sickness. Three kinds of sensory rearrangement involve visual/inertial conflicts, as exemplified by the 'cinerama sickness' in which the 'motion' carried by visual signals as one watches a chase scene is not correlated with appropriate vestibular signals. The remaining three types of rearrangement involve canal/otolith conflicts, such as those produced when an individual is rotated about an off-vertical axis or enters a zero-gravity environment. An interesting example of this second set of conflicts is positional alcoholic nystagmus after excessive ingestion of alcohol. Money and his colleagues have demonstrated that this phenomenon occurs as a result of changes in the specific gravity of the semicircular canal's cupula, which takes up alcohol from the blood more rapidly than the surrounding endolymph; consequently, the cupula tends to 'float' in the canal and the neural signals transmitted to the brain are the same as those ordinarily associated with rotation (as many who have overindulged can attest). Although this six-part classification of sensory rearrangements is incomplete, it represents a significant contribution toward systematizing the literature on motion sickness.

The practical and theoretical importance of the fact that individuals adapt to sensory rearrangements is explored in chapter 6. Given the opportunity to actively explore an altered inertial environment, disorientation and motion sickness gradually subside. If the exposure is sufficiently long — perhaps several days in a slowly rotating room — observers acquire the ability to move about without distress, although a high degree of drowsiness persists. On return to the normal environment, motion sickness reappears, and in those cases where polarity can

be observed, the signs are in the opposite direction from those seen during the initial exposure to the altered environment. Reason and Brand believe that "it can be argued that 'maladaptation sickness' is perhaps a more appropriate term than motion sickness; particularly since the condition can be reinstated by *removal* of the motion stimulus" (p. 145).

The central component of the theoretical model for adaptation to sensory rearrangement is a comparator unit. This component, which relates expected sensory patterns to those actually elicited, receives one of its inputs from a neural store of previous sensory patterns associated with a particular environment. Adaptation is the result of a change in the sensory pattern received by the comparator unit from the neural store. The authors suggest that overlearning may account for the observation that readaptation to a normal environment occurs more rapidly than the initial adaptation to sensory rearrangement.

Most of the research on sensory conflict to date has used visual rearrangements. Because several postural and eye-movement reflexes that are mediated by the vestibular system can be used to indicate the perceived orientation of an experimental animal, investigators may wish to consider the possibility of extending this approach to examine responses to inertial rearrangements.

Chapter 7 deals with the topic of the susceptibility to motion sickness. The most successful techniques for assessing susceptibility have used cross-coupled (Coriolis) acceleration, which is produced, for example, when an observer performs angular head motions while rotating about a vertical axis. Three dimensions are postulated to be important determinants of susceptibility: receptivity to accelerative stimuli, adaptability, and retentiveness. Correlates of susceptibility include age (highest between two and twelve) and sex (less frequently observed in men). Since it has been commonly believed that susceptibility to motion sickness is indicative of 'psychological weakness' or 'lack of moral fiber,' perhaps that susceptibility is accounted for at least partially by cultural sex-role demands.

Chapters 8 and 9 describe research on the prevention of motion sickness by drugs and preexposure adaptation, reviewing our current understanding of the neural and physiological mechanisms that mediate this response. These chapters will be of interest primarily to those who wish to acquire a more technical knowledge of this research.

Chapter 10, "Taking Stock," briefly reviews the basic points developed in the previous chapters and points out some of the limitations of those formulations. One major problem is not dealt with effectively: apparently, no reasonable hypothesis has been proposed to indicate why neurons activated by the vestibular system should be linked to a vomiting response. A second problem is the inability of Reason and Brand's theory to account for the failure of highly motivated individuals to exhibit symptoms of motion sickness, a finding which is perhaps related to the potentiating effect of focusing attention on internal states that was noted previously. Finally, from a neurophysiological/neuroanatomical view, the mechanisms by which the neural store of sensory patterns and the comparator are thought to function are uncomfortably vague.

The chapter ends with a proposed new definition: "Motion sickness is a self-inflicted maladaptation phenomenon, characterised primarily by the presence of pallor, cold sweating, nausea and vomiting, which occurs at the onset and cessation of conditions of sensory rearrangement when the pattern of inputs from the

vestibular system, other proprioceptors and vision is at variance with the stored patterns derived from recent transactions with the spatial environment" (p. 275).

As did Gibson's *The Senses Considered as Perceptual Systems* (1966) and Howard and Templeton's *Human Spatial Orientation* (1966), Reason and Brand's *Motion Sickness* ascribes to the vestibular system a central role in sensory psychology. Gibson's analysis implied that the vestibular system could be most readily understood by focusing on the perceptual goals that information from the vestibular receptors helps the individual achieve. Clearly, the vestibular system contributes information to a basic system for the perception of spatial orientation, a system that allows the individual to answer the complex questions 'Where am I?' and 'Where is the stimulus relative to me?' Gibson argued that we should examine sensory receptors not as sources of sensation but rather as extractors of information from the environment.

Howard and Templeton's book included a review of numerous psychophysical studies that analyzed vestibular contributions to perception of the vertical and to detection of motion. As Howard and Templeton made abundantly clear, the vestibular system is not alone in providing information on spatial orientation; skin receptors, eyes, ears, and perhaps even olfactory receptors also contribute. Further evidence that the vestibular system should be considered a component of a multimodal and basic system for the perception of spatial orientation derives from neurophysiological investigations. Fredrichson, Figge, Scheid, and Kornhuber (*Experimental Brain Research,* 1966, 2:318–327) located the primary vestibular receiving area in the somatosensory cortex of the rhesus monkey. This same somatosensory region receives input from joint receptors, and it appears that the neurons respond both to position of the limbs and to accelerative stimuli.

Reason and Brand's contribution is to demonstrate that vestibular receptors provide a major input to the basic perceptual system for spatial orientation. The importance of the vestibular input is illustrated by the severe disorientation that results when information from the vestibular receptors fails to coincide with information from other components of the basic perceptual system.

Given the apparently central role of the vestibular system in spatial orientation, why has this sensory system virtually been ignored by psychologists? That the vestibular system has been ignored can be deduced from a brief survey of introductory psychology textbooks, whose coverage of the vestibular system can range from nothing to two or three paragraphs. This paucity of coverage is manifested even in books devoted specifically to sensory psychology (e.g., Christman, *Sensory Experience,* 1971, and Scharf, *Experimental Sensory Psychology,* 1975), although Schiffman's *Sensation and Perception* (1976) departed from this trend with its chapter on the orienting system. I submit that the major reasons for our failure to embrace the vestibular system in sensory psychology include (a) an interaction between the initial research methodology of experimental psychology and the character of sensations associated with vestibular stimulation and (b) the fact that vestibular receptors are important contributors to the motor control system.

Wundt and his followers promulgated the view that the appropriate subject matter for psychology was the individual's internal experience: sensations, feelings, and so on. Given this methodological orientation, it is not surprising that

the vestibular system traditionally has been relegated to a minor role: the character of sensations associated with vestibular-receptor stimulation differs from that associated with other sensory systems, primarily in that the observer does not locate the sensation at the receptor.

It is an interesting speculation that our ability to localize sensations at particular receptors results in part from our ability to block stimulation of these receptors. We can hold our nose, blindfold our eyes, plug our ears, or bandage our skin, but we cannot cut off stimulation to our vestibular receptors by interrupting the path between the stimulus and the sensory receiver. In fact, the notion of a specific, localizable stimulus for the vestibular receptors is not meaningful (consider gravity). This speculation is supported by Melzack's observations on sensory deprivation in puppies.

While vestibular sensations are not localized at the receptors, it is incorrect to state (as do Reason and Brand, p. 85) that vestibular sensations only reach consciousness with overstimulation. Using a parallel (four-pole) swing that produces an oscillating linear acceleration and in which cues to motion at the eyes and ears are eliminated and tactile sensations masked by packing the observer in a styrofoam body mold, we have repeatedly demonstrated the exquisite ability of observers to judge linear accelerations of varying amplitudes (see Parker, *Journal of Experimental Psychology,* 1970, 84:96–104). When asked to describe their sensations in that swing, observers ordinarily report that they feel the motion either in the center of their heads or their stomachs.

The most influential twentieth-century neurophysiologist, Sherrington, recognized that the labyrinth was a source of sensation and struggled with the description of these sensations: "The labyrinthine receptors and their arcs give the animal its definite attitude to the external world. The muscular receptors give to the segment — e.g., the hind limb — a definite attitude less in reference to the external world than in reference to other segments, e.g., the rest of the animal. Our own sensations from the *labyrinth* refer to some extent, as said above, to this environment, that is, have some projected quality" (*The Integrative Action of the Nervous System,* 1906, p. 343). Unfortunately, Sherrington emphasized the role of the vestibular system in muscular regulation, and it is this emphasis that was subsequently pursued. Sherrington's influence can be seen in the inclusion of discussions of the vestibular system in chapters on motor control (e.g., Deutsch and Deutsch, *Physiological Psychology,* 1966).

In sum, it is time for psychologists to recognize that sensory events are embedded in a stable spatial framework that is the product of a basic perceptual system to which the vestibular system is a major contributor. Reason and Brand's book should aid this recognition. Perhaps the vestibular system and the center for the vomiting response were linked to help bring psychologists to their senses.

Reason and Brand argue that zero-gravity motion sickness results from a canal/otolith conflict and is basically the same as the sickness associated with ocean travel or exposure to the altered inertial environment of a slowly rotating room. Observations by Skylab astronauts militate against this argument. In spite of extensive preadaptation to canal/otolith conflict, the astronauts of Skylab 3 were greatly disturbed by the symptoms of motion sickness, and they said later they did not think the preadaptation procedures recommended by the vestibular

scientists had been useful. The Skylab 4 astronauts attempted to achieve pre-adaptation by acrobatic flight maneuvers instead, again without notable success.

In view of the apparent failure of models based on terrestrial observations to account completely for the Skylab reports, I have proposed a possible mechanism of vestibular activation by the zero-gravity environment that could either produce motion sickness alone or potentiate the response to other sensory conflicts. This proposal is based on four postulates. First, entry into the zero-gravity environment produces significant shifts in body fluids and results in increased fluid pressure in the cranial region. Second, as a result of this fluid shift, the fluid pressures in the perilymphatic and endolymphatic spaces are altered. Third, the altered fluid pressures of the inner ear result in changed activity of the vestibular neural pathway. Fourth, the resulting mismatch in neural signals from the two labyrinths, between labyrinth components, or between the labyrinths and other spatial-orientation receptors produces the symptoms of zero-gravity sickness, following Reason and Brand's view. Essentially what I have proposed is a Ménière's-disease model of zero-gravity sickness, which takes us full circle back to the beginning of the useful literature on motion sickness.

Future experiments will decide whether the canal/otolith view proposed by Reason and Brand or the altered-fluid-pressure hypothesis accounts for motion sickness in the zero-gravity environment.

Donald E. Parker, *Miami University*

Notes

Preparation of this review was supported by Contract NAS 9 14358 between Miama University and the National Aeronautics and Space Administration.

Contributors to Volume 88

References to book reviews and reports of meetings are italicized.

Baars, Bernard J. 467
Barash, David P. 195
Baumeister, Alfred A. 431
Benjafield, John 737
Bergum, Bruce O. *327*
Biederman, Irving 669
Birnbaum, Michael H. 417
Blake, Randolph *351*
Bootzin, Richard *325*
Bornstein, Marc H. 269
Bourne, Lyle E., Jr. 219, 499
Bowles, Lawrence J. 527
Brauner, Joel D. 105
Brown, Alan S. *185, 356*
Caprio, Michael 485
Carroll, John B. *172*
Cook, Thomas D. *343*
Crowder, Robert G. 115
Danks, Joseph H. 137
Duncan, Carl P. *331, 334, 353, 356, 562, 564*
Earhard, Bruce 385
Erwin, Donald E. 659
Flaherty, Charles F. 485
Freund, Joel S. 403, 695
Friedman, Alinda 601
Galbraith, Richard C. 515
Ghatala, Elizabeth S. 499
Gilbert, K. C. 431
Grant, David A. 147
Heller, F. P. 455
Herman, C. Peter *178*
Herzberger, Sharon, *562*
Hicks, Robert E. 303, 719
Hill, Winfred F. *190*
Hinman, Suki 403
Hollins, Mark 443
Houston, John P. 93, 507
Humphreys, Michael S. *351*
Izawa, Chizuko 681
Johnson, David M. 617
Kezwer, Ross 545
Kimmel, H. D. 553, 745
Kinsbourne, Marcel 719
Konewko, Michael 499
Lappin, Joseph S. *361*
Levin, Joel R. 499
Lewis, Donald J. 51
Lindauer, Martin S. *354*
Lit, Alfred 105
Locurto, Charles M. 281
McKelvie, Stuart J. 311
MacLeod, Colin M. 127
Maki, Judith 231
Massaro, Dominic W. *161*
Miller, George W. 303, 719
Moss, Leendert P. *332*
Motley, Michael T. 467
Nelson, Thomas O. 127
Newman, Slater E. 293
Nicholson, Lawrence R. 293
Noll, A. Michael 631
Olton, Robert M. 617
O'Neill, Maureen E. 535
Plutchik, Robert *557*
Rakover, Sam S. 241
Raye, Carol L. 645
Revelle, William *341, 350*
Richards, Larry G. 455, 707
Richardson, Jack 3
Roediger, Henry L., III 115
Royce, Joseph R. *332*
Runcie, Dennis 431
Runquist, Peggy A. 731
Runquist, Willard N. 231, 731
Salatas, H. 219
Sekuler, Robert 253
Silverstein, Charles I. 81
Slamecka, Norman J. 545
Spector, Amos 669
Staller, Joshua 253
Sutcliffe, Judith A. 535
Thomason, Terry L. 431
Tulving, Endel 535
Tzeng, Ovid J. L. 577
Walsh, John F. 281
Witte, Kenneth L. 695
Wong, Roderick 527
Wortman, Camille B. *330*
Wortman, Paul M. *328*
Yaroush, R. A. 499
Young, Robert *565*
Zechmeister, Eugene B. *344*

In memoriam

Wilfred John Brogden 147

Contents

W. B. Pavlik and Alexis C. Collier. Magnitude and schedule of reinforcement in rats' resistance to extinction — within subjects 195

Chok C. Hiew. Sequence effects in rule learning and conceptual generalization 207

William A. Ciszewski and Charles F. Flaherty. Failure of a reinstatement treatment to influence negative contrast 219

William C. Gordon. Similarities of recently acquired and reactivated memories in interference 231

Joseph A. Sgro. The use of relief and nonrelief from shock in the double alleyway 243

Drake R. Bradley and Heywood M. Petry. Organizational determinants of subjective countour: The subjective Necker cube 253

Pamela R. Sneed, David C. Brunts, and John H. Mueller. Instructions and grouped presentation in free recall 263

Leon Manelis. Frequency and meaningfulness in tachistoscopic word perception 269

Ruth H. Maki. Recall as a function of the relationship between the cues and the words to be remembered 281

Joseph J. Franchina, James R. Weeks, Jr., and Nanette Pais. Goal-box placements with initial nonreward and situational similarity in resistance to extinction 291

Ryan D. Tweney and Dan Swart. Experimental control of reaction times to negative and expletive sentences 299

Allen R. Dobbs. 'Forced' encoding but 'no' learning 309

Joan Regan. Strategies in learning for recall and recognition tests 313

H. D. Kimmel. Notes from "Pavlov's Wednesdays": Sensory preconditioning 319

Book reviews (Janet L. Mistler-Lachman, Carl P. Duncan, Ernest R. Hilgard, Joseph F. Lappin, Bennet B. Murdock, Jr., Eva D. Ferguson, Jack A. Adams, Robert Teghtsoonian, Willard Runquist, Daniel P. Keating, Kay Deaux, James K. Arima, L. Rowell Huesmann) 323

AMERICAN JOURNAL OF PSYCHOLOGY
June 1977, Vol. 90, No. 2, pp. 195–205

Magnitude and schedule of reinforcement in rats' resistance to extinction — within subjects

W. B. Pavlik and Alexis C. Collier
Virginia Polytechnic Institute and State University

Rats given acquisition and extinction training in discrete lever-pressing trials in lieu of the conventional runway were required to complete a fixed ratio of responses on a retractable lever to end each trial. Magnitude of reward (large or small) and schedule of reinforcement (continuous or partial) was signaled during each trial. Contrary to the classic effects of these variables between subjects, there was greater resistance to extinction with continuous than partial reinforcement (a reversed effect of partial reinforcement) at both magnitudes of reward; greater resistance to extinction with large than small magnitudes at both schedules; and no interaction of magnitude and schedule in resistance to extinction.

Conceptually, the analogy has been drawn between behavioral situations involving the rat in a runway and the rat in a lever-pressing apparatus, especially if the latter involves making a fixed number of responses on a retractable lever to complete each trial (e.g., Logan and Wagner, 1965; Platt, 1971). That the analogy is more than merely conceptual has been shown in a variety of experiments in which phenomena normally associated with the runway have been duplicated with discrete lever-pressing trials. Thus, Gonzalez, Bainbridge, and Bitterman (1966) showed that single-alternation response patterning (i.e., fast responding on reinforced trials and slow responding on nonreinforced trials), a phenomenon normally studied in the runway, could be duplicated in the situation with discrete lever-pressing trials.

More pertinent to the concerns of the present experiment, both Porter and Hug (1965), using rats and water reinforcement, and Bitgood and Platt (1971), using pigeons and food reinforcement, have shown that the conventional effect of partial reinforcement that is obtained between groups in runway situations can be duplicated in discrete-trials operant situations. Of considerable methodological importance in Bitgood and Platt's study was the inclusion of a group that was reinforced at the end

of each trial but which was required to make varying numbers of keypecks on different trials. Specifically, this group was required to make a given number of pecks on a particular trial, a number equal to the number of pecks made by the partially reinforced group between successive reinforcements at the corresponding point in training. Thus, this group was matched to the continuously reinforced group in terms of reinforcements per *trial* but was matched to the partially reinforced group in terms of reinforcements per *keypeck*. During extinction this third group performed like the continuously reinforced group and was significantly less resistant to extinction than the partially reinforced group, suggesting that the effect of partial reinforcement in a discrete-trials situation depends on reinforcements per trial rather than on reinforcements per unit response.

The purpose of the experiment reported here was to use that situation to examine the effect of partial reinforcement *within* subjects. Unlike the studies cited above, prior work on this problem in the runway situation does not provide any firm expectations as to the outcome. Various outcomes have been reported, the most common being that subjects given this kind of treatment fail to show any differential effect of schedules on resistance to extinction but are more resistant to extinction on both continuously and partially reinforced trials than is a control group given only continuous reinforcement and less resistant to extinction than a control group given only partial reinforcement (see Amsel, 1967). Such a result has been termed a 'generalized partial reinforcement effect' (see Mellgren and Dyck, 1972).

The present experiment also examined the effect of partial reinforcement within subjects as a function of the magnitude of the reward. The basis of this manipulation was the classic observation of Hulse (1958) and Wagner (1961) that the size of the effect of partial reinforcement between subjects in a runway was larger with larger rewards. Their results, viewed orthogonally, also indicated that the effects of the magnitude of the reward on resistance to extinction are schedule-dependent and asymmetrical. Given continuous reinforcement, a large reward led to large reductions in resistance to extinction; given partial reinforcement, a large reward led to moderate increases in resistance to extinction.

Recently, Pavlik and Collier (1973) have shown that the schedule-dependent effects of the magnitude of a reward on the resistance to extinction obtained between subjects could be utilized to eliminate the *reversed* effect of partial reinforcement that occurs within subjects in certain types of free-operant experiments (e.g., Pavlik and Carlton, 1965). Two separate groups of subjects were used, one trained with a larger and one trained with a smaller reward. Each group was trained and extin-

guished according to the procedures of Pavlik and Carlton (1965). As was predicted from the findings of Hulse (1958) and Wagner (1961), the group with the small reward showed greater resistance to extinction after continuous than partial reinforcement (the reversed effect), but no effect of schedule occurred in the group with the large reward. Compared with the small reward, the large reward failed to produce the reversed effect, in most part because of a sizeable decrement in resistance to extinction after continuous reinforcement.

Although the results of the Pavlik and Collier study suggest that the magnitude of the reward has the same schedule-dependent effects on resistance to extinction within subjects in a free-operant context as it does between subjects in a discrete-trials context, they leave unanswered the question of whether these schedule-dependent effects are replicable when both the magnitude of reward and the schedule of reinforcement are simultaneously varied within subjects. This question formed an additional basis for the present experiment.

Thus, unlike prior studies, the present experiment examined possible interactive effects of schedule of reinforcement and magnitude of reward when both variables were simultaneously manipulated within subjects on a discrete-trials basis. Specifically, each subject in the present experiment experienced trials under four different conditions that represented the factorial combination of continuous or partial reinforcement schedules with large or small rewards. Different combinations of discriminatory stimuli signaled which schedule and which magnitude of reinforcement were in effect on each trial.

METHOD

Subjects

The subjects were seven Long-Evans male hooded rats, approximately 120 days old at the start of the experiment, from the local departmental colony. Four rats had been used previously in a free-operant experiment (Pavlik and Collier, 1973), and three were experimentally naive. (As no differences in the experimental performance of naive and experienced rats were apparent, this difference in preexperimental history will be ignored below.)

Apparatus

A BRS test chamber equipped with two retractable levers was used. The chamber contained dim cue lights mounted over the levers, relatively bright house lights mounted at the rear of the chamber, and a pellet dispenser programmed to dispense either one or four 45-mg Noyes pellets. A BRS Audio Generator (model AU-902) was programmed to deliver 81-db inputs to the experimental chamber, each input being either white noise or a clicking sound

(approximately 10 clicks per sec). The experimental procedures and contingencies were controlled by associated programming equipment.

Procedure

All rats were maintained on ad lib water in the home cage and a total daily food ration of 12 g, consisting of powdered Purina Chow adjusted for the weight of the Noyes pellets consumed in the test chamber. After six days of accommodation to the deprivation schedule and handling, pretraining was begun. Pretraining consisted of magazine and bar training, which continued until all rats were performing in a relatively rapid and stable manner on an FR-10 reinforcement schedule. During pretraining, the lever at the right was continuously present, as were the lighted cue lights and the white noise. Reinforcements during pretraining consisted of one 45-mg Noyes pellet.

The experiment per se was run in two cycles. Each cycle consisted of both an acquisition phase and an extinction phase. Both acquisition and extinction phases consisted of daily sessions of discrete trials. Specifically, each trial started with the insertion of one of the levers into the test chamber and the presentation of a compound, visual and auditory, stimulus. The trial continued until the tenth lever press, which resulted in the retraction of the lever, termination of the stimulus, presentation of the appropriate reinforcement (if any), and the initiation of a 30-sec intertrial interval. On all the trials, latency was measured from the insertion of the lever to completion of the FR-10, and was recorded to the nearest .1 sec by a printout counter.

Cycle 1

During cycle 1, for all subjects, the lever at the left was programmed to produce large, four-pellet, reinforcements, and the lever at the right produced small, one-pellet, reinforcements. Accompanying insertion of the lever, either of two compound stimuli was presented. One compound, consisting of lighted cue lights over the levers and the clicking sound, signaled continuous reinforcement. The other compound, consisting of lighted house lights and the white noise, signaled partial reinforcement. By combining lever position, left or right, with the particular compound, a total of four different types of trials were possible: large reward, continuous reinforcement (LC); large reward, partial reinforcement (LP); small reward, continuous reinforcement (SC); and small reward, partial reinforcement (SP). During the intertrial interval, the white noise and lighted cue lights were present in the chamber.

Each rat received a total of 20 trials during each acquisition session. Trials were arranged in blocks of four such that each of the four types of trials (LC, LP, SC, and SP) occurred once in each block. Further, each type of trial occurred equally often in each ordinal position within blocks and followed each other type of trial equally often, but never followed itself (i.e., the same type of trial never occurred twice in succession). Thus, five trials of each type occurred during each session. Two LP and two SP trials were reinforced during each session, a 40% reinforcement schedule, and the pattern of reinforcement was varied from session to session. Four different sequences of 20 trials meeting the above criteria were used. They were varied across sessions in a quasi-random pattern.

During cycle 1, each rat received 23 daily acquisition sessions, for a total of 115 trials of each of the four types. On the day after the last acquisition session, extinction was begun. A total of four extinction sessions were given, one per day. The first extinction session consisted of 24 trials. The first four trials, one of each type, were all reinforced. The remaining 20 trials were nonreinforced, as were all 20 trials on each of the final three extinction sessions. During extinction, a 60-sec response limit was invoked. If subjects had not completed the FR-10 within 60 sec, the trial was terminated automatically and a latency of 60 sec was recorded.

Cycle 2

Six weeks after the completion of cycle 1, during which time the rats were maintained in their home cages on the daily food ration of 12 g but otherwise had no contact with any of the experimental procedures, cycle 2 was begun. Cycle 2 was undertaken to assess the effects of the confounding, during cycle 1, of particular stimulus and particular reinforcement conditions. Since all rats had had exactly the same stimulus condition paired with exactly the same reinforcement parameter during cycle 1, any differences in performance could be attributed either to differences among the reinforcement conditions or to differences among the stimulus conditions. During cycle 2, for all rats, the correlations between particular stimulus conditions and particular reinforcement conditions were reversed as completely as possible. Specifically, whereas lever position previously signaled magnitude of reward, it now signaled schedule of reinforcement, the left lever signaling partial reinforcement and the right lever signaling continuous reinforcement. The compounds that previously had signaled schedule now signaled magnitude, the compound of lighted cue lights and clicking now signaling the small, one-pellet, reward and the compound of lighted house lights and white noise now signaling the large, four-pellet, reward. Thus, the combination of lever position and stimulus compound that previously had signaled the most favorable reinforcement contingency, LC, now signaled the least favorable contingency, SP, and the combination that previously had signaled SP now signaled LC. Similar exchanges occurred between the LP and SC conditions.

During cycle 2, all rats received 25 daily acquistion sessions of 20 trials each, followed by five daily extinction sessions. All procedures, except those noted here, were identical with those of cycle 1.

RESULTS

One rat died between cycle 1 and cycle 2. Its data are included in the results for cycle 1. For all rats, a mean latency was computed for each acquisition and extinction session for each of the four types of trials (LC, LP, SC, and SP). The principal results are depicted in Figure 1, which indicates group means across acquisition and extinction sessions, and in Figure 2, which indicates both group and individual subject means collapsed across extinction sessions for both cycles. (The rat that died is not

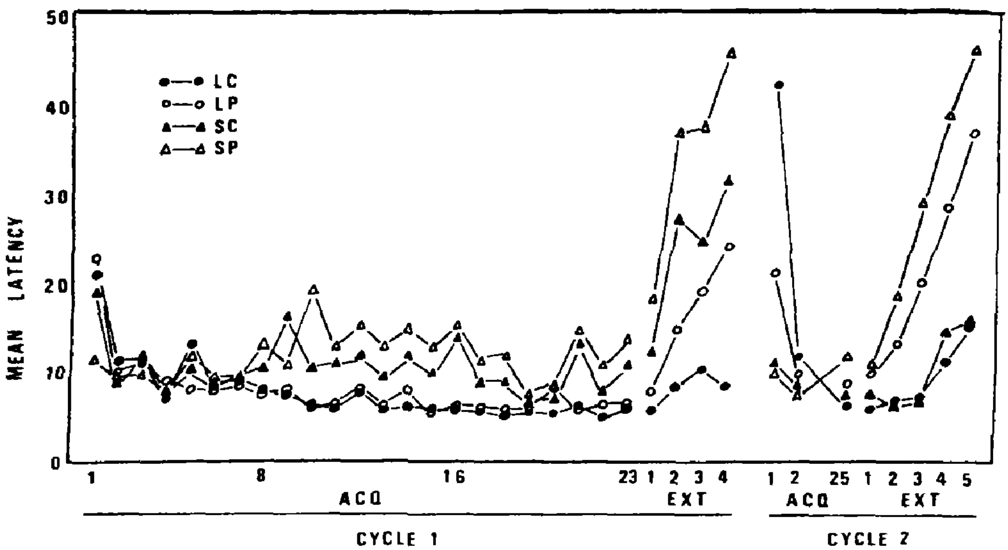


Figure 1. Group mean latencies (in seconds) across acquisition (ACQ) and extinction (EXT) sessions during the two experimental cycles for each of the four types of trials, LC, LP, SC, and SP

included among the individual subject means for cycle 1, although its data are included in the group means.)

Cycle 1

Reference to Figure 1 indicates that the rats were showing some discrimination among the four different types of trials during the acquisition phase. Mean latencies computed over the last five acquisition sessions for each type of trial were subjected to an analysis of variance, which yielded, using a pooled error term, a significant effect of magnitude of reward [$F(1, 18) = 24.9$, $p < .001$], with the larger reward producing shorter latencies; an insignificant effect of schedule of reinforcement [$F(1, 18) = 2.27$, $p > .10$], with continuous reinforcement tending to produce shorter latencies; and the absence of any interaction of the two [$F < 1.0$].

Extinction performance during cycle 1, as is evident in Figure 1, showed clear differences among the four types of trials, differences which increased as extinction progressed. Contrary to the between-subjects results of Hulse (1958) and of Wagner (1961), greater resistance to extinction was associated with continuous than with partial reinforcement (the reversed effect) at both magnitudes of reward, and with a large than with

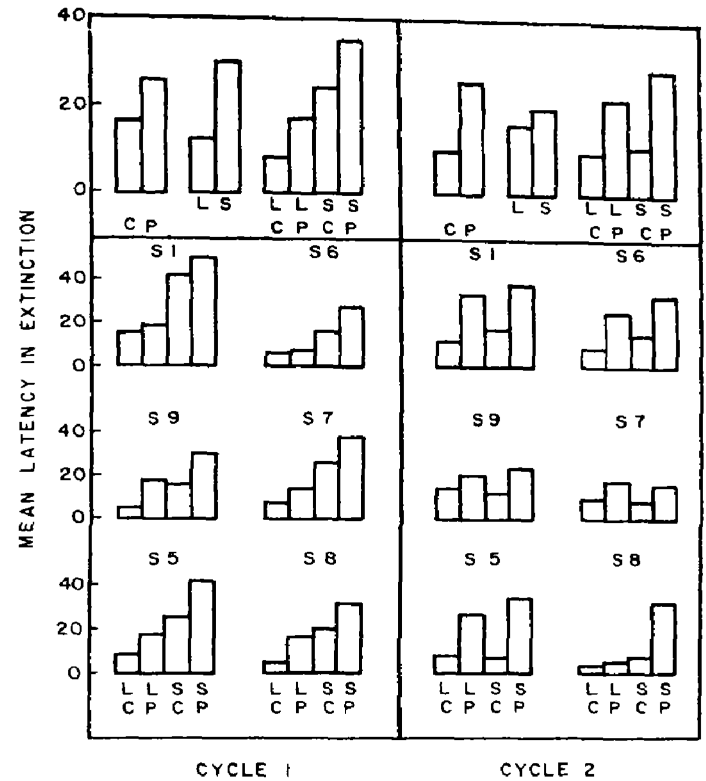


Figure 2. *Mean latencies (in seconds) during extinction, collapsed across sessions, associated with continuous reinforcement* (C), *partial reinforcement* (P), *large reward* (L), *small reward* (S), *and the factorial combinations: group means are shown in the top two panels, individual-subject means in the bottom two panels*

a small reward at both schedules. Analysis of the mean latencies shown in Figure 2 yielded significant [$p < .001$] effects of both schedule [$F(1, 18) = 31.99$] and magnitude [$F(1, 18) = 100.12$]; the interaction was not significant [$F < 1.0$]. It should be noted in Figure 2 that each individual subject yielded a pattern of results (lower portion of the figure) consistent with the pattern shown by the group means (upper portion of the figure).

Cycle 2

The first results of interest in cycle 2 are those of the first acquisition session, where some large differences among the four types of trials are evident in Figure 1. It should be noted that the ordering of the latencies

on this day is exactly the reverse of the ordering that obtained during the late acquisition and extinction phases of cycle 1. It will be recalled that the procedures of cycle 2 involved the complete reversal of the relations obtaining between particular stimulus conditions and particular reinforcement parameters (see above). These differences thus suggest that even after four extinction sessions and six weeks of no contact with the experimental situation, the rats remembered the original correlations of cues with reinforcement parameters that obtained during the acquisition phase of cycle 1.

By the end of the acquisition phase of cycle 2, the rats were again showing evidence of discrimination among the four types of trials, with shorter latencies occurring with continuous reinforcement and/or a large reward. Mean latencies over the last five acquisition sessions were subjected to analysis of variance, yielding a significant effect of schedule [$F(1, 15) = 8.12, p < .02$], a marginal effect of magnitude [$F(1, 15) = 2.77, p < .20$], and the absence of any interaction [$F < 1.0$].

The extinction results of cycle 2 were similar to those of cycle 1 (see Figure 1). Again, greater resistance to extinction was associated with continuous than with partial reinforcement (the reversed effect) at both magnitudes of reward, and with a large reward than with a small reward at both schedules. The mean latencies shown in Figure 2 were subjected to analysis of variance. The effect of schedule was highly significant [$F(1, 15) = 43.14, p < .001$]. The effect of magnitude was barely nonsignificant [$F(1, 15) = 3.22, p < .10$]. The interaction [$F(1, 15) = 1.66$] was clearly not significant.

DISCUSSION

The present experiment examined resistance to extinction as a joint function of schedule and magnitude of reinforcement purely within subjects. The principal findings were that (a) resistance to extinction was greater with continuous than with partial reinforcement at both magnitudes of reward, (b) resistance to extinction was greater with a large reward than a small reward at both schedules, and (c) magnitude and schedule of reinforcement did not interact in affecting resistance to extinction. All three findings are, of course, different from the analogous results of investigations of the same variables purely between subjects (e.g., Hulse, 1958; Wagner, 1961). It also is clear that the present findings, particularly the reversed effect of partial reinforcement, conflict with what one normally would expect to be the predictions of most current theories of instrumental appetitive conditioning. A closer examina-

tion of the implications of some of these theories suggests, however, that the reversed effect that was obtained may not be theoretically as anomalous as it first appears.

While we have been unable to detect any way in which Amsel's frustration theory (Amsel, 1967) can predict the reversed effect within subjects, the sequential theory of the conventional effect between subjects (Capaldi, 1967), as extended by Mellgren and Dyck (1972) and Rudy, Homzie, Cox, Graeber, and Carter (1970), does predict both the reversed effect within subjects under certain conditions and the conventional effect within subjects under other conditions. The conditions determining which outcome occurs involve the way in which transitions from nonreward to reward (N-R) during acquisition are correlated with shifts between the discriminanda used to signal the continuous and partial reinforcement schedules. Such correlations of N-R and schedule transitions were not controlled by the particular counterbalancing techniques used to generate trial sequences in the present experiment, leaving indeterminate the relevance of the present results to sequential theory. However, further work in our laboratory replicating Mellgren and Dyck's (1972) procedures, but with the discrete lever-pressing trials of the present study, has consistently produced robust reversed effects within subjects, effects which appear impervious to the loci of N-R transitions (Collier, Steil, and Pavlik, 1975).

Note should be made of the differences in the results of cycle 1 and cycle 2 of the present experiment (see Figure 2). The effects generally were larger and more clear-cut in cycle 1. Further, while magnitude of reward had a larger effect than did the schedule of reinforcement during cycle 1, the opposite was true during cycle 2. Since the chief procedural difference between the cycles was the complete reversal of the relationships between specific cues and specific reinforcement parameters (i.e., in some sense, cycle 2 constituted a discrimination-reversal task for the subjects), we might expect the results during cycle 2 to be diminished by proactive interference from cycle 1. Also, it should be noted that the more effective variable during cycle 1, magnitude, and the more effective variable during cycle 2, schedule, were both signaled by lever position, while the less effective variable in each cycle was signaled by the compound auditory and visual stimulus. A relative superiority of a spatial cue to a visual cue in the discrimination of reinforcement conditions was found by Spear and Spitzner (1969) in a rather different context.

Finally, while the results of the present experiment are virtually the opposite of those typically obtained between subjects with nondiscriminative instrumental appetitive conditioning, such results are not unknown in the field of animal learning. Greater resistance to extinction both with

continuous reinforcement and with larger rewards has been reported between subjects in discrimination settings (e.g., Pavlik and Reynolds, 1964; Spear and Spitzner, 1969), within subjects in free-operant studies with multiple schedules (e.g., Pavlik and Carlton, 1965; Shettleworth and Nevin, 1965), and between subjects in studies of Pavlovian conditioning (e.g., Scheuer, 1969; Wagner, Siegel, Thomas, and Ellison, 1964). It is worth noting that these instances of the 'reversed' effect involve cases where there was some explicit pairing of an external cue with a particular condition of reinforcement and that the resistance to extinction shown in the presence of a cue was greater the more favorable the condition of reinforcement paired with that cue during acquisition. In their accounts of the conventional effect of partial reinforcement between subjects, both Amsel (1967) and Capaldi (1967) assume that the greater resistance to extinction shown by partially reinforced subjects stems from their experiencing, during extinction, internal cues (s_f for Amsel and S^N for Capaldi) that had been paired reliably during acquisition with reinforcement.

Thus, conceptual congruity of the results of the present experiment with more conventional findings may be claimed by asserting that both represent the general case of the resistance to extinction being greater the more favorable the reinforcement conditions during acquisition with which were paired the functional stimuli operating from moment to moment during extinction. The problem remains, of course, of being able to specify the functional stimuli operating at any given moment during extinction.

Notes

Requests for offprints should be sent to W. B. Pavlik, Department of Psychology, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061. Alexis C. Collier is now at Ohio State University. Received for publication March 10, 1976; revision, June 7, 1976.

References

Amsel, A. 1967. Partial reinforcement effects on vigor and persistence: Advances in frustration theory derived from a variety of within-subjects experiments. In *The psychology of learning and motivation,* ed. K. W. Spence and J. T. Spence, vol. 1, New York: Academic Press.

Bitgood, S. E., and Platt, J. R. 1971. A discrete-trials PREE in an operant situation. *Psychonomic Science* 23:17–19.

Capaldi, E. J. 1967. A sequential hypothesis of instrumental learning. In *The psychology of learning and motivation,* vol. 1, ed. K. W. Spence and J. T. Spence. New York: Academic Press.

Collier, A. C., Steil, D. A., and Pavlik, W. B. 1975. Within-subjects PREs: The absence of N-R transition effects. Paper presented at the meeting of the Midwestern Psychological Association in Chicago in May.

Gonzalez, R. C., Bainbridge, P., and Bitterman, M. E. 1966. Discrete-trials lever pressing in the rat as a function of pattern of reinforcement, effortfulness of response, and amount of reward. *Journal of Comparative and Physiological Psychology* 61:110–122.

Hulse, S. H. 1958. Amount and percentage of reinforcement and duration of goal confinement in conditioning and extinction. *Journal of Experimental Psychology* 56:48–57.

Logan, F. A., and Wagner, A. R. 1965. *Reward and punishment*. Boston: Allyn and Bacon.

Mellgren, R. L., and Dyck, D. G. 1972. Partial reinforcement effect, reverse partial reinforcement effect, and generalized partial reinforcement effect within subjects. *Journal of Experimental Psychology* 92:339–346.

Pavlik, W. B., and Carlton, P. L. 1965. A reversed partial-reinforcement effect. *Journal of Experimental Psychology* 70:417–423.

Pavlik, W. B., and Collier, A. C. 1973. Reinforcer magnitude effects on a within-subjects reversed PRE. *Bulletin of the Psychonomic Society* 2:233–234.

Pavlik, W. B., and Reynolds, W. F. 1964. Conventional and reversed partial reinforcement effects in selective learning. *Psychonomic Science* 1:155–156.

Platt, J. R. 1971. Discrete-trials and their relation to free-behavior situations. In *Essays in neobehaviorism: A memorial volume to Kenneth W. Spence*, ed. H. H. Kendler and J. T. Spence. New York: Appleton-Century-Crofts.

Porter, J. J., and Hug, J. J. 1965. Effects of number and percentage of rewarded trials on the acquisition and extinction of lever pressing using a discrete-trial procedure. *Journal of Experimental Psychology* 70:575–579.

Rudy, J. W., Homzie, M. J., Cox, R. W., Graeber, R. C., and Carter, E. N. 1970. Effects of sequential manipulations in the within-subjects partial reinforcement experiment with rats. *Journal of Comparative and Physiological Psychology* 72:105–115.

Scheuer, C. 1969. Resistance to extinction of the CER as a function of shock-reinforcement training schedules. *Psychonomic Science* 17:181–182.

Shettleworth, S., and Nevin, J. A. 1965. Relative rate of response and relative magnitude of reinforcement in multiple schedules. *Journal of the Experimental Analysis of Behavior* 8:199–202.

Spear, N. E., and Spitzner, J. H. 1969. Extinction effects of percentage reinforcement and reinforcer magnitude within subjects: The influence of discrimination task. *Psychonomic Monograph Supplements* 3:19–24.

Wagner, A. R. 1961. Effects of amount and percentage of reinforcement and number of acquisition trials on conditioning and extinction. *Journal of Experimental Psychology* 63:234–242.

Wagner, A. R., Siegel, S., Thomas, E., and Ellison, G. D. 1964. Reinforcement history and the extinction of a conditioned salivary response. *Journal of Comparative and Physiological Psychology* 58:354–358.

AMERICAN JOURNAL OF PSYCHOLOGY
June 1977, Vol. 90, No. 2, pp. 207–218

Sequence effects in rule learning and conceptual generalization

Chok C. Hiew
University of New Brunswick, Canada

Subjects were given 12 training problems based on four primary bidimensional rules, the problems mixed so that the rule changed after every problem (to facilitate adoption of a truth-table strategy; condition M) or systematically blocked by rule (condition S). Their next tasks were to learn an unpracticed complementary rule, to discover a tridimensional rule, and to identify a rule based on novel and meaningful stimuli. The findings show that condition M facilitated the subjects' ability to abstract, generalize, and extend a higher-order strategy. They suggest the importance of the sequence as well as the variety of rules and imply a hierarchical model of concept learning.

A series of experiments (e.g., Bourne, 1970; Bourne and Guy, 1968) demonstrated that practice at learning bidimensional rules produces considerable positive intrarule transfer as well as interrule transfer. With appropriate experience, initial differences between rules disappeared, and performance came close to being errorless. Analysis of these data suggested that subjects learn a strategy that simplifies their rule learning, and Bourne (1967) argued that many subjects, after solving a series of problems based on the four primary rules (see Table 1), acquire the same general strategy, a strategy best described as an intuitive version of the logical truth table.

Specifically, the subject is able to collapse and code the entire stimulus population into four classes based on the presence and absence of the two relevant attributes — into classes that correspond to the true-true (TT), true-false (TF), false-true (FT), and false-false (FF) values of a conventional truth table. After learning to code all stimuli into these four classes, the subject has only to find their category assignments to discover the rule. Thus, a subject using the optimal truth-table strategy could discover any bidimensional rule with at most one error per stimulus class. This is supported by Bourne's (1970) finding that the proportion of subjects achieving optimal solutions increased as practice (number of prob-

Table 1. Assignments of stimulus classes to response categories (+ and −) under the four primary and four complementary bidimensional rules

Stimulus class	Truth-table values	Stimulus set	Bidimensional rules							
			Cj	Cj*	Dj	Dj*	Cd	Cd*	Bd	Bd*
RS	TT	RS	+	−	+	−	+	−	+	−
$R\bar{S}$	TF	RTr, RC	−	+	+	−	−	+	−	+
$\bar{R}S$	FT	GS, BS	−	+	+	−	+	−	−	+
$\overline{RS}$	FF	GTr, GC, BTR, BC	−	+	−	+	+	−	+	−

Note: T represents true (or present); F, false (or absent); R, red; G, green; B, blue; S, square; Tr, triangle, and C, circle. The bidimensional rules, reading from right to left, are conjunctive, conjunctive complement, disjunctive, disjunctive complement, conditional, conditional complement, biconditional, and biconditional complement.

lems) increased. By the thirteenth problem, 83% of the subjects had reached optimal solution.

The present study was designed to answer two questions: first, whether acquisition of the truth-table strategy can be facilitated, and second, the extent to which sophisticated subjects' knowledge and skills are transferable to new and novel tasks. In Bourne's (1970) study, the four primary rules were presented systematically with the three problems, each rule being administered in blocks. Mastery of each rule was achieved methodically and rapidly, because the same rule was repeated sequentially. The present study was concerned with whether the subjects could learn both the rules and a general strategy for learning other, unknown rules if the problems were randomly mixed in such a way that the primary rule changed after each problem and could not be used to predict the next problem.

A number of considerations suggested that such a mixed condition (condition M) would be better than a systematic condition (condition S) in the acquisition of the truth-table approach to solving problems. Early literature on simple discrimination learning by animals (e.g., Levine, Levinson, and Harlow, 1959) and humans (e.g., Morrisett and Hovland, 1959; Clarke and Cooper, 1966) has shown that training which stresses both learning within problems and generalization across problems produces the most interproblem transfer to new stimuli. Thus, training under condition M was expected to more enhance generalization across problems than was training under condition S, because the former presents the different rules successively, without focusing at length on each specific rule.

Several studies (e.g., Haygood, Sandlin, Yoder, and Dodd, 1969) have

confirmed that the contiguity principle (Underwood, 1952) is quite a general phenomenon in concept learning. Increased contiguity of instances of the same concept was found to aid performance. Bourne (1970) theorized that just as presenting instances of a rule usually leads to learning of the specific rule, presenting a variety of rules in a series of problems could lead to the discovery of a superordinate concept or strategy. In other words, each rule can be regarded as an instance of the general concept and the entire series of problems as a single task. It may not be unreasonable to assume that condition M offers greater contiguity of instances of the superordinate truth-table concept but less contiguity of instances of subordinate rules than condition S. Hence, it was expected that training under condition M would produce a better acquisition of the strategy.

To investigate the extent to which knowledge of the conceptual strategy was transferable, three tests of generalization were given after training: transfer to a new, unknown bidimensional rule that was presented as for training (*complementary* task, see Table 1); transfer to a more complex rule that was based on three relevant attributes and required coding the stimuli into eight classes, all stimuli presented simultaneously (*tridimensional* task); transfer to a bidimensional rule that used novel stimuli constructed from drawings of faces and was presented as a 'real life' problem (*meaningful* task). The expectation was that training under condition M, having produced more truth-table strategists than condition S, would lead to better performance in each transfer task. A control group (condition C) was given no training and tested on only the transfer tasks.

METHOD

Subjects and design

The subjects were 108 undergraduates from the University of Colorado, solicited through advertisements in the daily student newspaper and paid to participate in the experiment. Of them, 36 were randomly assigned to each of three groups: conditions M, S, and C. All subjects in conditions M and S solved 15 successive problems. The first 12 were rule-learning problems and constituted the training. For each subject, 3 of the 12 training problems were based on each of the four primary bidimensional rules. Under condition M, the 3 problems based on each rule were randomized, the rule changing after every problem. Under condition S, the 3 problems based on each rule were grouped into blocks and administered successively so that the rule changed after problems 3, 6, and 9. Under both conditions, the rules themselves were counterbalanced, with all possible orders being used. A total of 24 different orders was generated and 12 additional orders repeated for the 36 subjects used. Besides being constrained by an equal number of orders starting with each of the four primary rules, the 36 orders further used all rules equally and also had an equal number of orders ending with the same rule. Additionally,

under condition M, in each series of 4 problems, all four primary rules were represented.

For the generalization tasks, all subjects solved three problems. Problem 13 always involved a complementary rule. For the control subjects (condition C), this was the first problem encountered. For each experimental subject, the given complementary rule selected was unlike the complement of the rule in problem 12. Equal numbers of subjects under each condition (9) were given each of the four complementary rules. The final two problems, 14 and 15, were the tridimensional task and the meaningful task, their order of presentation counterbalanced. The rule learned in the tridimensional task was based on a compound and/or combination of the three relevant attributes, while the conditional rule was used in the meaningful task.

Attributes and stimuli

For the first 13 problems, the stimuli were rear projected on a translucent screen. The stimuli were geometric patterns and varied on four dimensions, each with three levels: color (red, yellow, or black), form (square, triangle, or hexagon), number of figures (one, two, or three), and size (small, medium, or large). The correct classification was completely determined by the relevant attributes (e.g., red and square) on two of these dimensions, plus the rule. Each problem had a different pair of attributes. To speed learning, only three of the four dimensions varied in any problem, with one attribute from the fourth dimension always constant. Thirteen pairs of relevant attributes were randomly chosen from the 54 possibilities and assigned unsystematically to the 13 problems. For the complementary task, problem 13, the pair of relevant attributes chosen did not vary for the four complementary rules.

The stimuli for the tridimensional task were geometric patterns constructed of masonite cutouts pasted on a 14-by-14-in. cardboard. The stimuli had four bilevel dimensions: form (triangle or square), color (blue or green), size (large or small), and number of dots (one or two). All 16 different patterns were displayed on a chart and laid out randomly. They could be coded into eight classes: one class (corresponding to a TTT value) with all three relevant attributes, three classes (TTF, TFT, and FTT) with two of the relevant attributes, three classes (TFF, FTF, and FFT) with one relevant attribute, and finally, one class (FFF) with none of the relevant attributes present.

For the meaningful task, each stimulus was a drawing of the outline of a pig's face made from white index cards pasted on a black background. The faces varied on three bilevel dimensions: eyes (slant or crossed), ears (pointed or flopped), and nose (round or flattened). All eight stimuli were pasted on a 30-by-11-in. cardboard. Four drawings were marked with three positive and one negative sign so that the conditional rule was exemplified. These four marked stimuli and the remaining four unmarked stimuli could be separately exposed to the subject by covering them up appropriately with strips of cardboard.

Tasks and procedure

All subjects were instructed to find a relation or rule between two given attributes that would allow errorless classification of the geometric patterns. Stimulus patterns were presented and response requirements were made as described by Bourne (1970). In addition, the subject's response latencies were measured (in sec) from the presentation of the stimulus to his depression of

the response switch. After a problem was solved, the subject had to describe the rule without any prompting by the experimenter. Under condition S, the experimenter indicated whether the next concept was based on the same or a different (unnamed) rule. Under condition M, subjects were reminded that the rule was different from the one in the problem just solved.

For problem 13, the complementary task, the subject was told that he had to learn a new unpracticed rule using an identical procedure. In the case of the control subjects, condition C, the task was their very first problem, and they were given instructions similar to those given the experimental subjects before training.

For the tridimensional task, a chart displaying all of the patterns was placed in front of the subject. The three relevant attributes were given to the subject (small, blue, square), and he was told to select only eight patterns to solve the problem. After each selection, the experimenter gave the appropriate categorization on a marker placed on the pattern chosen. When the subject had made eight choices, he then described the rule. The subject was cautioned to select only the most informative patterns without any redundant information on them. He was given paper and pencil and as much time as he wanted to plan his eight choices.

For the meaningful task, the subject was presented with a chart showing four stimuli (uncategorized). The experimenter explained that the subject had to figure out the rule that a pig-breeder used to sort out "good" pigs from "poor" ones. The subject was given the two relevant attributes (slant eyes, pointed ears) upon which the breeder based his rule. After the experimenter finished explaining, the four stimuli were covered up, and four new categorized samples were exposed to the subject for him to examine for about 2 min. When the subject indicated he was ready, these four samples were covered up, and the initial four uncategorized stimuli were presented to the subject for classification. Finally, the subject described the rule used.

RESULTS

Acquisition of truth-table strategy

Three measures are reported: mean trials to solution averaged across the four stimulus classes, mean response latencies in seconds for each stimulus class, and the subjects' verbal descriptions of the rules. A description was scored as exemplifying a coded or truth-table strategy only if the presence and absence (or "not") of the relevant attributes was used. A description that simply listed attributes or dimensions was scored as uncoded.

The overall main effect of mixed or systematic training (condition M or S) showed no reliable differences for mean trials or response latencies. The interaction of condition by problem, however, was significant [$F(11, 770) = 3.82, p < .001$] for mean trials (see Figure 1). Initially, in problem 1, there was no difference in performance between conditions. However, in problems 4, 7, and 10, each of which involved a new rule for

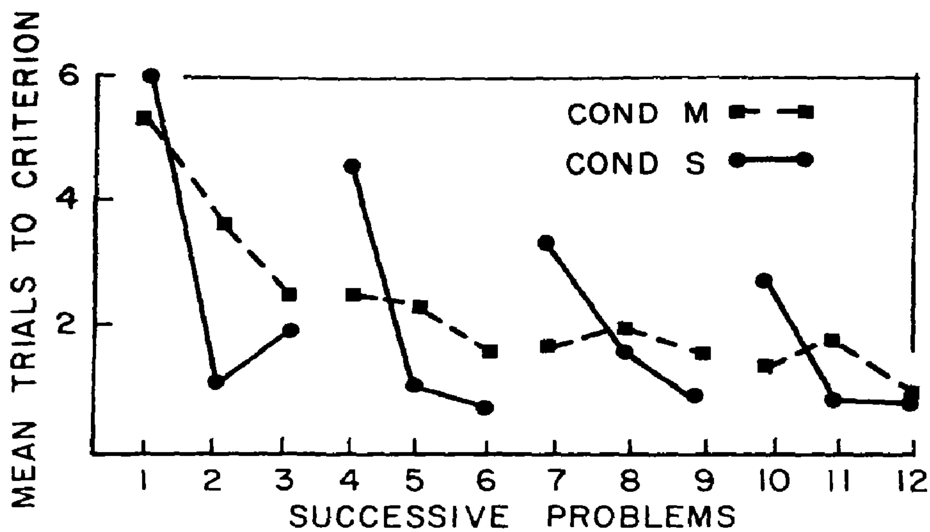


Figure 1. Mean trials to criterion for each stimulus class summed across the four stimulus classes

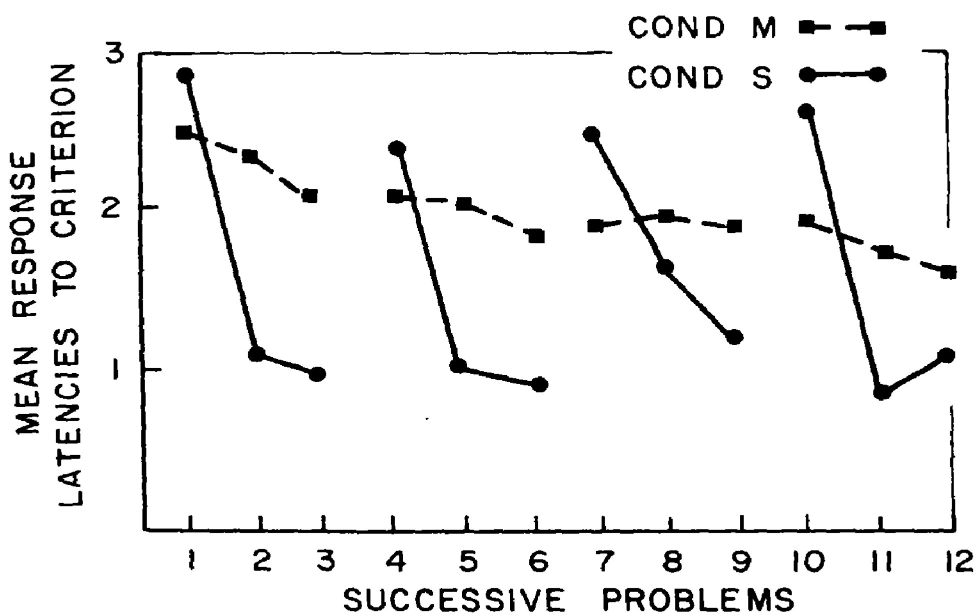


Figure 2. Mean response latencies to criterion for each stimulus class summed across the four stimulus classes across successive problems

subjects under condition S and a change of rule (but not necessarily to a new rule) for subjects under condition M, the latter showed a greater improvement than the former [$F(1, 70) = 6.62, p < .025$]. In one sense, interrule transfer under condition M exceeded that under condition S. Predictably, performance under condition S on the second and third problem with each rule was superior to that under condition M [$F(1, 70) = 13.72, p < .001$]. This indicated the considerable amount of interproblem transfer under condition S. Yet by the last training problem (problem 12), despite their many changes of rule under condition M, these subjects showed the same level of performance as those under condition S.

Figure 2 shows the response latencies (in sec). There was a significant interaction of condition by problem [$F(11, 770) = 2.99, p < .001$]. The subjects under condition M showed a fairly constant response latency across successive problems, while those under condition S showed a rapid decline within each block of problems. The suggestion is that subjects under condition S may not have been attending to the stimuli for long enough after each presentation of a new rule to abstract a general strategy; apparently, they were responding to the stimuli by following the particular rule just learned, and not learning enough about the stimuli to be able to code them in a general way.

Optimal solution

The percentage of subjects who gave optimal solutions — that is, who solved the problem with no more than one instance of each of the stimulus classes — was computed. Condition M produced numerically more optimal solutions than condition S on problem 4 (9, or 25%, and 6, or 16%, respectively), problem 7 (17, or 47%, and 12, or 33%, respectively), and problem 10 (15, or 42%, and 12, or 33%, respectively). Considering only the harder primary rules (Cd and Bd) during problems 4, 7, and 10, condition M produced a total of 14 optimal solutions (attributed to 13 subjects) compared to 7 in condition S (attributed to 7 subjects). These findings suggested that condition M tended to produce more subjects with the theoretically optimal strategy than condition S.

An expectation was that subjects who gave coded descriptions of the rules would be closer to optimal solution. The finding was that under condition M, 95% of the subjects who solved optimally gave coded verbal descriptions; under condition S, the statistic was 85%. Moreover, there was a strong negative correlation between proportions of coded descriptions and performance during training [$r = -.60, p < .001$]. Figure 3 shows the percentage of subjects who gave coded descriptions. Analysis

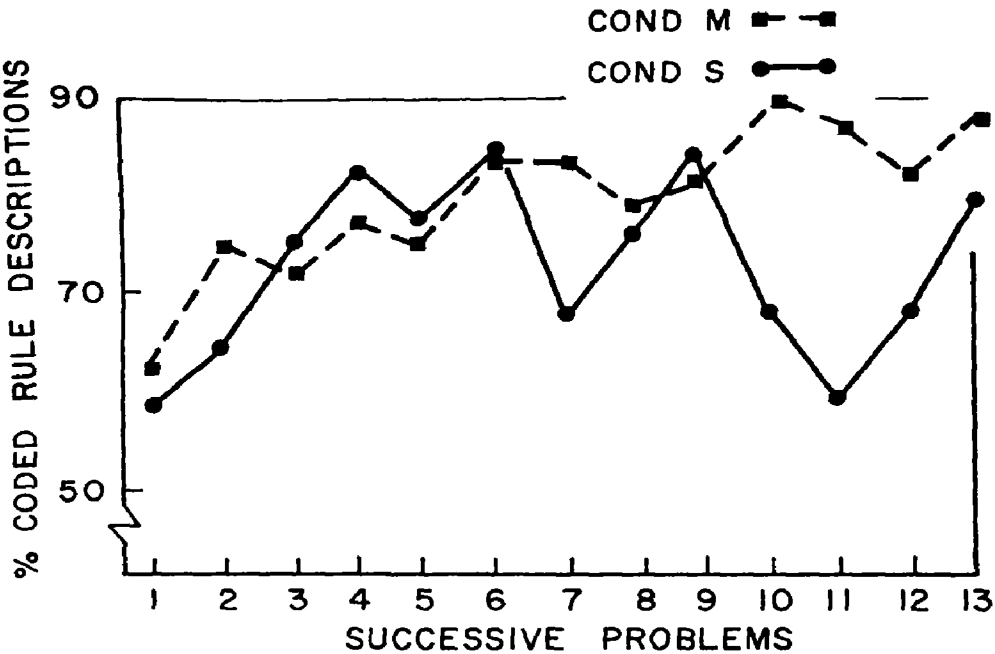


Figure 3. Percentage of subjects reporting coded descriptions across problems

indicated a reliable difference during training, with a mean of 79% under condition M and 69% under condition S [$F(1, 22) = 5.89, p < .025$].

Inspection of Figure 3 shows a gradual but consistent increase in percentage of coded descriptions under condition M, while improvement under condition S tended to be erratic across problems. This indicates that subjects under condition M were gradually learning and using an optimal strategy. In the case of condition S, the poor correspondence between the subjects' ease of solving rules in each block of problems and the percentage of subjects reporting coded descriptions suggests that these subjects may not have been resorting to a truth-table approach.

Complementary task

As predicted, there was a significant main effect of condition [$F(2, 96) = 16.25, p < .001$], with mean trials of 1.05, 2.40, and 7.03 under conditions M, S, and C respectively. Orthogonal comparisons indicated that conditions M and S were reliably different from condition C [$F(1, 96) = 31.21, p < .001$] and that condition M was not significantly different from condition S. The nonsignificant difference between the two experimental conditions is misleading, as it was essentially due to the large error variance contributed by naive subjects under condition C. This is strongly

supported by the finding that the ratio of their mean square error variance was highly significant (204.7/28.8). Accordingly, a 2 × 4 ANOVA of experimental conditions by complementary rules was done, and it showed that condition M was, in fact, reliably superior to condition S [$F(1, 64) = 6.38$, $p < .01$]. Furthermore, there was a significant difference between rules [$F(3, 64) = 6.14$, $p < .001$] and a significant interaction between condition and rule [$F(3, 64) = 2.82$, $p < .04$]. Simple main-effect analysis of the interaction showed that there were no appreciable differences in performance among the four complementary rules under condition M, but vast differences under condition S [$F(3, 64) = 8.40$, $p < .0001$]. It appears that subjects under condition M were not only superior at solving complementary rules but could apply the strategy with practically equal facility to a variety of new rules, whereas subjects under condition S could not.

Response latencies of the subjects showed a sginificant effect of condition [$F(2, 96) = 5.89$, $p < .005$], with mean latencies of 1.67, 2.00, and 3.05 sec under conditions M, S, and C respectively. Orthogonal comparisons show that conditions M and S were reliably different from condition C [$F(1, 96) = 11.14$, $p < .005$], while conditions M and S did not differ. Neither the main effect of rule nor the interaction of condition by rule reached an acceptable level of significance. To evaluate possible sources of rule difficulty, latencies on stimuli comprising the four stimulus classes were evaluated. Under condition M, there were no differences in latencies among rules nor was there an interaction of rule by stimulus class. Under condition S, the variability in responding to the stimuli was indicated by the reliable difference in latencies among rules [$F(3, 32) = 2.92$, $p < .05$], as well as by a significant interaction of rule by stimulus class [$F(9, 96) = 2.91$, $p < .01$]. In addition, these subjects took more time to respond to stimuli in the more difficult Cd* and Bd* rules than the simpler Cj* and Dj* rules. These findings indicated that subjects under condition S were unable to code all stimuli within the four complementary rules with equal facility.

Optimal solutions were given by 20 subjects (55.6%) under condition M compared to 16 (44.4%) under condition S. Also, 86% of the rule descriptions under condition M were coded, as compared to 78% under condition S.

Tridimensional task

Analysis of variance indicated a reliable difference among the three conditions [$F(2, 105) = 3.38$, $p < .05$]. Orthogonal comparisons, however, showed that while performance under condition M was reliably

superior to that under condition S [$F(1, 105) = 6.01, p < .025$], the performance of the subjects under condition S was no better than that under condition C. The number of subjects who correctly selected all eight stimuli was 18 (50%) under condition M, 9 (25%) under condition S, and 8 (22.2%) under condition C. These findings showed a significantly greater ability on the part of subjects under condition M to generalize a bidimensional strategy to a tridimensional task. On the other hand, subjects under condition S, despite their training, did not perform with greater success than subjects under condition C.

An ideal strategy for solving the tridimensional task would be to select one stimulus from each of the eight stimulus classes depending on the presence and absence of the three relevant attributes. It was found that the single largest source of omissions came from the three stimulus classes with only one relevant attribute present (TFF, FTF, and FFT). However, subjects under condition M were more successful at selecting these stimuli than those under conditions S and C. Chi-square tests indicated a reliable difference between conditions M and S [$\chi^2(1) = 10.28, p < .01$] and between conditions M and C [$\chi^2(1) = 3.76, p < .06$]. Surprisingly, the subjects under condition S performed numerically less well than those under condition C. Training under condition M, being more effective for the acquisition of a truth-table strategy, led to a greater ability to discriminate between, and to code for, these three stimulus classes than training under condition S. Presumably, the training under condition S brought about a set to solve the problems quickly, leading to poorer attention to details or differences between stimuli.

Meaningful task

Analysis failed to indicate significant differences in the number of stimuli correctly classified under the three conditions. The subjects under each condition performed equally well (mean of 3 out of a total of 4 classifications). The number of subjects who classified all stimuli correctly were 18, 15, and 13 under conditions M, S, and C respectively. There was a lack of coded descriptions and the number of such descriptions again did not vary between the conditions. The task required subjects to identify the Cd rule, but few subjects gave descriptions resembling that rule, even subjects who correctly classified all stimuli. Most subjects reported idiosyncratic rules listing attributes which included irrelevant dimensions. Apparently, the task was simple enough, even the control subjects performing extremely well on it, that it did not require the experimental subjects to learn the less salient truth-table strategy.

DISCUSSION

The hypothesis that increased variety of rules aids subsequent transfer has previously been affirmed (Bourne, 1970; Bourne and Guy, 1968). The present research expands this finding to show that in addition to variety, the sequence of rules also affects transfer. Specifically, mixing the sequence of problems (condition M) so that the rule changes on consecutive problems leads to better transfer than systematically blocking problems by rule (condition S). The findings suggest that such superior performance was facilitated by the adoption of an intuitive version of a truth-table strategy. More than this, the implication is that it makes sense to consider the strategy as a superordinate concept and the series of problems as instances of the concept. This assumption calls for a hierarchical model of conceptual behavior (Bourne, 1970).

After solving a series of problems with bidimensional rules, most subjects can give an efficient description of the rules. More important, subjects can also learn a more general rule (the strategy) on how to solve problems based on a variety of bidimensional rules. The subjects have acquired a higher-order rule that relates lower-order rules within the conceptual system of rules. The rules are acquired in a hierarchical fashion, with higher-order concepts being transformed from, and built on, lower-order concepts. With the generalization tasks, the prediction that subjects under condition M would have a greater ability to generalize to new problems was confirmed (except for the meaningful task, where all subjects, including the controls, performed equally well). Under condition M, the strategy generalized with virtually equal facility to all four complementary rules, and the mean trials to solution of problem 13 approximated the theoretical optimal solution of at most one error per stimulus class. Moreover, these subjects took the same amount of time to respond to each stimulus regardless of its class or rule. Another finding is that these subjects' knowledge efficiently extended to solve a complex problem involving tridimensional rules. These findings are consistent with the notion that the strategy is equally applicable to the whole conceptual system of logical rules.

Training that has been made difficult, either by increasing intratask interference (Battig, 1972) or by presenting complex stimuli (Clarke and Cooper, 1966), can have a facilitative effect on retention and on transfer to new tasks. The present experiment indicates that how concepts are sequenced strongly affects their utilization in new tasks. Under condition M, changing the rule after solution of each problem produced a more gradual interproblem transfer and a slower response latency. However,

the greater complexity encountered during this acquisition was compensated by a greater flexibility of application of the general strategy to meet the demands of new tasks. These findings point to a negative aspect of learning tasks based solely on the contiguity of similar basic concepts. The superior performance under condition M suggests that for best results, there is a need for concurrent learning that emphasizes the fundamental strategy or higher-order concept underlying the subordinate concepts as well.

Notes

This report is based on a dissertation submitted by the author to the Graduate School of the University of Colorado in partial fulfillment of the requirements for the Ph.D. The author thanks his research advisor, L. E. Bourne, Jr., for his critical advice. A version of this report was presented at the meeting of the Rocky Mountain Psychological Association in Salt Lake City in May, 1975. Requests for offprints should be sent to C. C. Hiew, Department of Psychology, University of New Brunswick, Fredericton, New Brunswick, Canada. Received for publication April 28, 1976.

References

Battig, W. F. 1972. Intratask interference as a source of facilitation in transfer and retention. In *Topics in learning and performance,* ed. R. F. Thompson and J. F. Voss. New York: Academic Press.

Bourne, L. E., Jr. 1967. Learning and utilization of conceptual rules. In *Concepts and the structure of memory,* ed. B. Kleinmuntz. New York: Wiley.

Bourne, L. E., Jr. 1970. Knowing and using concepts. *Psychological Review* 77:546–556.

Bourne, L. E., Jr., and Guy, D. E. 1968. Learning conceptual rules: I. Some interrule transfer effects. *Journal of Experimental Psychology* 76:423–429.

Clarke, A. M., and Cooper, G. M. 1966. Transfer in category learning of young children: Its relation to task complexity and overlearning. *British Journal of Psychology* 57:121–128.

Haygood, R. C., Sandlin, J., Yoder, D. L., and Dodd, D. H. 1969. Instance contiguity in disjunctive concept learning. *Journal of Experimental Psychology* 81:605–607.

Levine, M., Levinson, B., and Harlow, H. F. 1959. Trials per problem as variable in the acquisition of discrimination learning set. *Journal of Comparative and Physiological Psychology* 52:396–398.

Morrisett, L., and Hovland, C. I. 1959. A comparison of three varieties of training in human problem solving. *Journal of Experimental Psychology* 58:52–55.

Underwood, B. J. 1952. An orientation for research on thinking. *Psychological Review* 59:209–220.

AMERICAN JOURNAL OF PSYCHOLOGY
June 1977, Vol. 90, No. 2, pp. 219-229

Failure of a reinstatement treatment to influence negative contrast

William A. Ciszewski and Charles F. Flaherty
Rutgers–The State University

In two experiments, rats were shifted from a 32% to a 4% sucrose solution after retention intervals of 1, 17, or 32 days. Negative contrast in lick rate was obtained after the intervals of 1 and 17 days, but not after 32 days. A reinstatement treatment consisting of periodic exposures to the preshift sucrose solution during the retention interval did not influence the degree of negative contrast obtained after the interval.

Rats shifted from a 32% sucrose solution to a 4% sucrose solution lick less after the shift than animals trained with only the 4% solution (Vogel, Mikulka, and Spear, 1968; Flaherty, Capobianco, and Hamilton, 1973). This response decrement appears to be directly analogous to the negative contrast found in instrumental paradigms (Crespi, 1942; DiLollo and Beez, 1966).

Previous research has shown that the degree of negative contrast after a shift in the quantity or quality of reward is a function of the length of the retention interval between shift phases. Specifically, for instrumental responding in a runway, degree of negative contrast has been found to vary systematically with retention intervals of 1 to 68 days (Gleitman and Steinman, 1964; Gonzales, Fernhoff, and David, 1973). Similarly, for consummatory responding, negative contrast has been obtained after a shift in the concentration of sucrose solutions with retention intervals of 1, 4, and 5 days between shift phases, but not with retention intervals of 17 or 32 days (Gordon, Flaherty, and Riley, 1973; Flaherty et al., 1973).

The occurrence of negative contrast implies that rats remember some attributes of the preshift reward (Spear, 1967), and conversely, the failure to obtain contrast suggests the possibility that rats forget some elements of that original training. If this is the case, procedures designed to maintain or arouse memory of attributes of the original reward should lead to the occurrence of contrast over longer retention intervals than would other-

wise be expected. For example, it has been demonstrated under other paradigms that periodic exposure to elements of the situation for original training is enough to maintain a learned response at a high level, but not enough to produce an effect in animals that had not received the original training (Campbell and Jaynes, 1966). This exposure to elements of original training has been termed reinstatement, and it has been found that such treatment enhances retention in both appetitive and aversive instrumental tasks (Campbell and Jaynes, 1966; Silvestri, Rohrbaugh, and Riccio, 1970; Greenfield and Riccio, 1972; Campbell and Jaynes, 1969; Gatti, Pais, and Weeks, 1975).

In a previous study of negative contrast in consummatory responding (Gordon et al., 1973), reinstatement was attempted after a retention interval of 32 days. The reinstatement treatment consisted of allowing the experimental rats ten licks of the original sucrose solution (either 32% or 4%) just before the postshift testing with the 4% solution. This treatment was ineffective: neither experimental nor control animals exhibited negative contrast during postshift testing. This failure to obtain reinstatement may have been due to several factors. The exposure to the original sucrose solutions may have been too brief to arouse the specific elements of preshift training relevant to negative contrast (see Campbell and Spear, 1972); or too long a period of time may have been allowed to elapse before the reinstatement treatment was administered (Flaherty et al., 1973); or the specific context in which the reinstatement treatment was administered (one different from preshift and postshift testing) may have detracted from the animals' retrieval capabilities (Spear, 1973; Gatti et al., 1975). The following experiments were conducted to assess the effects of a reinstatement treatment on negative contrast under more favorable conditions.

EXPERIMENT I

The present experiment replicated the preshift procedures described by Gordon et al. (1973) with these modifications in the reinstatement treatment: (a) exposure to the preshift reward during each session of the reinstatement treatment was increased and administered periodically during the retention interval, and (b) the reinstatement treatment was given in the original training apparatus. The reinstatement treatment consisted of a 45-sec exposure to the preshift concentration of sucrose solution every 6 days during the retention interval (unpublished data has indicated that contrast effects are maintained at least as long as 10 days).

METHOD

Subjects and apparatus

Thirty-two adult male Sprague-Dawley rats were used as subjects. Half of them were naive and had been bred in the Rutgers colony from Carworth-purchased parents. The other half were purchased from Carworth Laboratories: six were naive, and ten had been used several months earlier in an unrelated study using sucrose. Subjects were individually housed and maintained; the lighting cycle was 12 hours on, 12 hours off. All subjects were reduced to 85% of their (295–375 gm) ad lib weight and maintained at that level throughout the experiment by once-a-day feeding. Water was always available in the home cage.

Testing was conducted in three identical Plexiglas chambers described in detail in Flaherty et al. (1973). Licks were recorded by means of a contact relay circuit between the wire-mesh floor of the chamber and the graduated drinking tubes mounted on one wall of the chamber.

Procedure

Subjects were assigned to one of six groups defined by the factorial combination of preshift sucrose concentration (4%; 32%) and retention condition (1 day, no reinstatement; 32 days, no reinstatement; 32 days, reinstatement). Assignments were made randomly with the restriction that approximately equal numbers of experienced animals appear in all groups.

During the preshift days, the procedure was identical for all retention conditions. Each subject was placed in the test chamber and allowed a 5-min access to the appropriate concentration of sucrose solution. After each training period, the subject was weighed, returned to its home cage and fed its daily ration approximately half an hour later. There were ten preshift days.

In the postshift stage, subjects trained on a 32% concentration were shifted down to 4%, whereas subjects originally trained on a 4% concentration were maintained at 4%. Other aspects of this procedure were identical to the preshift phase. The shift for rats in the 1-day condition occurred the day after the last preshift day. These rats were run for a total of four postshift days. For the animals in the 32-day conditions, the first postshift test occurred 32 days after the last preshift day. These rats were maintained at 85% of their ad lib weight and were weighed every fifth day during the retention interval. Once postshift testing was initiated, the procedure was identical to that used with the 1-day groups.

During the 32-day retention period, the reinstatement and no-reinstatement groups were treated differently. The treatment employed with the former consisted of placing each subject in the original test chamber and allowing it a 45-sec access to the appropriate preshift concentration. This was done every sixth day after the last day of preshift training and on the day immediately preceding the postshift test. During this phase, the number of licks was recorded for each 15-sec exposure. During this period, rats in the no-reinstatement groups were just weighed and handled every fifth day. The concentrations of sucrose solution were prepared by weight from commercial-grade cane sugar and tap water.

RESULTS

One animal under the 32-day no-reinstatement condition was dropped from the experiment for failure to lick the solution in the preshift period. The mean number of licks for all groups in the preshift and postshift phases of the experiment is presented in Figure 1. Analysis of terminal preshift lick rates (mean of last three days) revealed an arithmetic, but not statistically reliable, effect of the concentration of the solution [for the 1-day group, $F(1, 6) = 3.04$, $p > .05$; for the 32-day groups, $F < 1.00$]. There was also *no* effect of reinstatement treatment, a pseudovariable at this point [$F(1, 19) = 1.47$, $p > .05$], nor was there an interaction of reinstatement treatment by concentration [$F < 1$].

In the postshift phase, rats in the 1-day group that were shifted from a

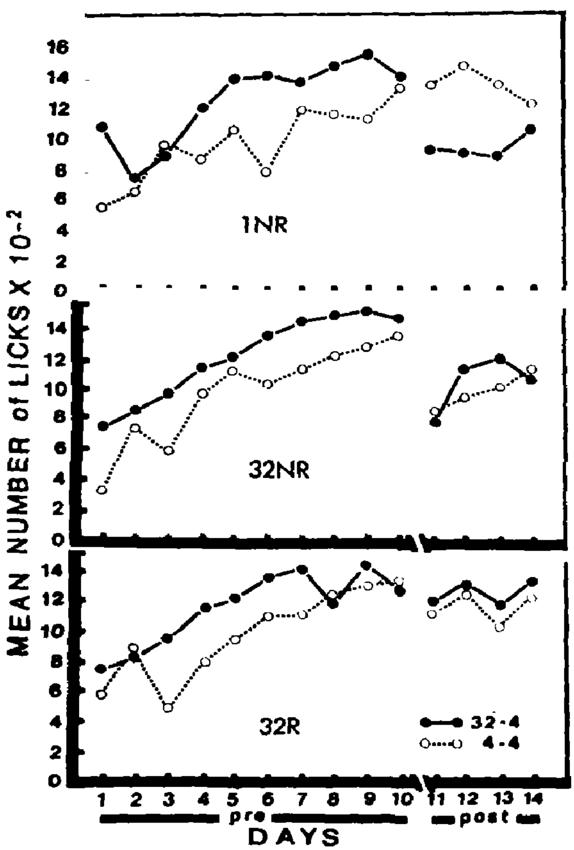


Figure 1. Mean lick rate during preshift and postshift phases as a function of concentration of sucrose solution (32% or 4%), retention interval, and reinstatement (R) or no reinstatement (NR)

32% to 4% concentration showed a decrease in licks per session to a point significantly below the unshifted 4% controls [$F(1, 6) = 6.76$, $p < .05$]. These data are presented in the top panel of Figure 1.

During the retention interval, the mean lick rates over the six 45-sec reinstatement treatments were 206.7 and 173.43 for the animals with 32% and 4% concentrations respectively. An analysis of variance performed on these data indicated that the difference was not statistically reliable [$F(1, 10) = 1.94$, $p > .05$].

Rats shifted from a 32% to 4% solution after a 32-day retention interval failed to exhibit contrast [$F < 1.00$]. That is, the shifted animals licked at the same rate as the unshifted controls. In addition, no reliable difference was found between the lick rates in the reinstatement and no-reinstatement groups [$F(1, 8) = 3.34$, $p > .05$], nor was an interaction of shift by reinstatement treatment obtained [$F < 1.00$]. None of the possible interactions between days and reinstatement treatment or shift approached statistical reliability.

However, an analysis of variance comparing the last preshift day with the first postshift day for rats in the 32-day groups indicated a reliable interaction of reinstatement treatment by shift phase [$F(1, 19) = 9.13$, $p < .01$]. Subsequent analysis with the Fisher *LSD* procedure [$p = .05$] indicated that the animals given the reinstatement treatment licked at an overall higher rate on the first postshift day than did the animals without such treatment.

DISCUSSION

Shifting rats from 32% to 4% sucrose solutions resulted in a negative contrast when a retention interval of 1 day, but not of 32 days, separated shift phases. These results are consistent with those found by Gordon et al. (1973) and imply a loss of the attributes of original training over the longer interval (see Spear, 1973).

In this experiment, we proceeded on the premise that if enough elements of the situation for the original training were presented during the retention interval, then the loss of negative contrast should be reduced. Such a result would be expected on the basis of other studies of reinstatement (e.g., Campbell and Jaynes, 1966; Shubat and Whitehouse, 1968; Silvestri et al., 1970). The present results indicate, however, that the reinstatement treatment did not reduce the loss of contrast, although it did have some effect on retention in that there was less decrement in overall lick rate for the animals given the reinstatement treatment.

Since it has been shown that the degree of forgetting is inversely related

to the duration of the reinstatement treatment (Campbell and Jaynes, 1969), the inability of the present treatment to reduce the loss of contrast may have been due to the parameters selected despite the fact that reinstatement exposure was increased substantially from that in the study by Gordon et al. (1973). Perhaps a longer reinstatement exposure and a shorter interval between treatment sessions would make retrieval of attributes of the preshift reward more likely. In order to assess this possibility, a second experiment was performed.

EXPERIMENT II

The purpose of the second experiment was to evaluate the effect of longer and more frequent sessions of the reinstatement treatment over a shorter retention interval. In this experiment, a 68-sec reinstatement exposure to the preshift solution was administered every 4 days, itself a retention interval that has been clearly shown *not* to lead to a loss in contrast (Flaherty et al., 1973). Total reinstatement exposure in Experiment II (272 sec) was equivalent to that in Experiment I. In addition, some animals received the reinstatement treatment only (no acquisition training).

METHOD

Subjects and apparatus

Forty-two naive adult male Sprague-Dawley rats, purchased from Blue Spruce Farms, were used as subjects. All animals were reduced to 85% of their (290–320 gm) ad lib weight and maintained at that level throughout the experiment by once-a-day feeding. In all other respects, Experiment II was identical to Experiment I, except that the lighting cycle was 8 hours off, 16 hours on. The apparatus was the same as in Experiment I.

Procedure

The procedure in Experiment II was generally the same as in Experiment I. Subjects were randomly assigned to one of six groups defined by the combination of preshift sucrose concentration (4%; 32%) and condition (17 days, reinstatement; 17 days, no reinstatement; reinstatement only).

Preshift procedure under all but the last of these conditions was identical to Experiment I. Subjects with reinstatement only were just weighed and handled each day during the preshift phase.

During the retention interval, subjects in the 17-day reinstatement and the reinstatement-only groups were placed into the original test chamber and allowed a 68-sec access to the appropriate preshift sucrose solution every fourth day following the last preshift day. The 17-day no-reinstatement rats were handled and weighed every fourth day during this phase. Postshift procedure was identical to Experiment I but continued for three days only.

RESULTS

The mean number of licks for each group in the preshift and postshift phases of the second experiment is presented in Figure 2.

In the preshift phase, the animals with a 32% concentration licked at a higher rate than those with a 4% concentration. This was verified statistically by performing an analysis of variance on the mean lick rates over the last three preshift days. The analysis indicated a significant effect of concentration [$F(1, 24) = 22.75, p < .001$], but no effect [$F < 1.00$] of reinstatement treatment, a pseudovariable at this point.

During the retention interval, data were lost from the second session of the reinstatement treatment due to equipment failure. Analysis of the data from the last two sessions of the reinstatement treatment indicated

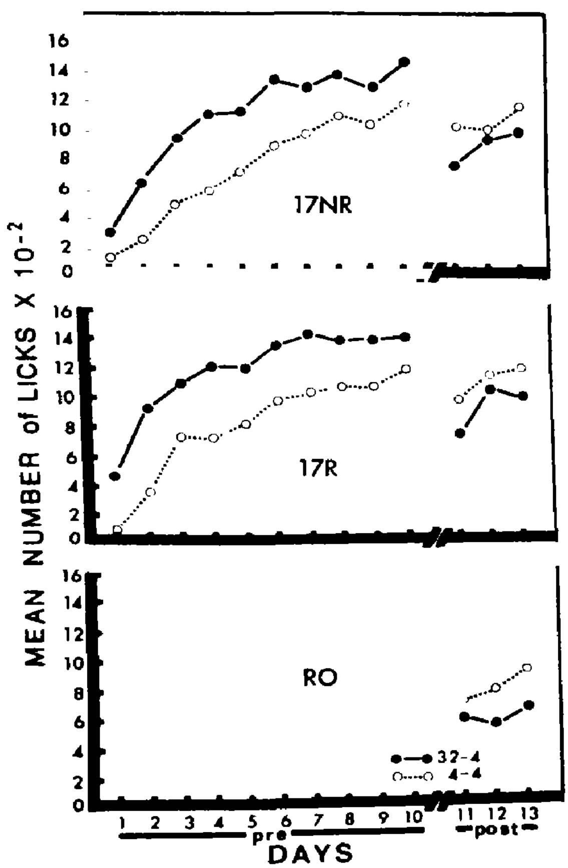


Figure 2. Mean lick rate during preshift and postshift phases as a function of concentration of sucrose solution (32% or 4%), and reinstatement (R), no reinstatement (NR), or reinstatement only (RO)

the following. The mean lick rates for subjects with a 32% concentration were 335.2 and 280.1 for acquisition training and reinstatement only respectively. For subjects with a 4% concentration, the mean lick rates were 221.28 and 107.14 respectively. An analysis of variance performed on these data indicated a reliable effect of the concentration of the sucrose solution [$F(1, 24) = 31.16, p < .0001$] and a reliable effect of acquisition training on lick rate [$F(1, 24) = 89.5, p < .0001$]. There was no reliable interaction of concentration by reinstatement treatment [$F(1, 24) = 3.78, p > .05$].

In the postshift phase, rats shifted from 32% to 4% concentration after a 17-day retention interval showed a decrease in licks per session to a point reliably below the unshifted 4% controls [$F(1, 24) = 4.82, p < .05$]. There was, however, no difference in lick rate between the animals with and without the reinstatement treatments [$F < 1$], nor was there an interaction of shift by reinstatement treatment [$F < 1$]. A significant overall increase in lick rate was obtained over the three days of the postshift phase [$F(2, 48) = 12.20, p < .01$], but no interactions were obtained with any of the treatment conditions.

The shifted animals that received reinstatement only showed a lick rate below the unshifted 4% controls; however, this difference did not reach an acceptable level of statistical significance [$F(1, 12) = 4.02, p < .07$]. The analysis also indicated that neither the effect of days [$F(2, 24) = 3.21, p < .07$] nor the interaction of shift by days [$F < 1.00$] was reliable.

DISCUSSION

In Experiment II, shifting rats from 32% to 4% concentrations of sucrose solution resulted in negative contrast when a retention interval of 17 days separated shift phases. The reinstatement treatment failed to influence the degree or duration of contrast obtained. This is despite the fact that for those animals that received the reinstatement treatment only (no acquisition training), there was marginally reliable contrast in the postshift period. It should be noted that although contrast was obtained after the 17-day retention interval in this experiment (but not in the study by Flaherty et al., 1973), the degree of contrast was not large and there was certainly 'room' for the reinstatement treatment to produce a larger contrast effect. This may be seen by comparing the contrast ratios (lick rate of shifted animals divided by lick rate of unshifted animals) obtained in the present experiment with contrast ratios obtained with other retention intervals. The contrast ratios of the animals with and

without the reinstatement treatment during the 17-day retention interval were .75 and .73 respectively (in comparison to a contrast ratio of approximately .87 in the experiment by Flaherty et al., 1973). However, contrast ratios of about .50, or less, are usually obtained with a 1-day retention interval, and a contrast ratio of .59 was obtained with a 4-day retention interval (Flaherty et al., 1973). Thus, there appears to be a gradient of forgetting on which the 17-day retention interval is at a point of marginally reliable contrast.

Considering both Experiments I and II as well as the study by Gordon et al. (1973), it is clear that the satisfaction of the operational requirements of reinstatement, as defined by earlier investigators (Campbell and Jaynes, 1966; Spear, 1973), did not serve to alter the apparent rate of forgetting of attributes of the preshift reward over long and moderate retention intervals. It seems unlikely that this failure was due to parametric considerations, since the control group in Experiment II that received reinstatement only fell just short of demonstrating reliable negative contrast. Increasing reinstatement exposure under the present paradigm would doubtlessly produce contrast effects equivalent to acquisition training.

Comparing the present experiments with earlier demonstrations of reinstatement, several possible reasons for the differences in results suggest themselves. For example, many studies that have found reinstatement have involved either fear conditioning or presenting the nonreinforced stimulus from an appetitive discrimination task (Campbell and Jaynes, 1966; Greenfield and Riccio, 1972; Campbell and Randall, 1976). Perhaps stimuli associated with aversiveness or with response inhibition are more effective *cues* for retrieval (see Campbell and Spear, 1972). Under the paradigm for successive negative contrast, there are no cues differentially associated with the quality of the reward or with aversiveness, nor is there occasion for response inhibition until the postshift phase of the experiment (see Black, 1968). Thus, if differential stimuli and aversiveness or inhibition are important for the reinstatement treatment to be effective, then the same effectiveness would not be expected under the paradigm for successive negative contrast. A determination of the importance of these factors, as well as a difference between instrumental and consummatory behavior, might be made by examining the effects of reinstatement under other paradigms for the study of contrast effects (e.g., Flaherty, Riley, and Spear, 1973).

Whatever the case, the present experiments have indicated that when consummatory negative contrast fails to occur after long retention intervals, a reinstatement treatment does not facilitate retrieval of attributes

of the preshift reward. When moderate negative contrast does occur after intermediate retention intervals, a reinstatement treatment fails to enhance or prolong the contrast. Thus, these experiments indicate that the effectiveness of reinstatement may not hold equally well in all learning situations.

Notes

Supported by a grant from the Rutgers Research Council and a Biological Sciences Support Grant, both to the second author. Offprint requests should be addressed to Charles F. Flaherty, Psychology Department, Busch Campus, Rutgers University, New Brunswick, New Jersey 08903. Received for publication March 26, 1976.

References

Black, R. W. 1968. Shifts in magnitude of reward and contrast effects in instrumental and selective learning: A reinterpretation. *Psychological Review* 75:114–126.

Campbell, B. A., and Jaynes, J. 1966. Reinstatement. *Psychological Review* 73: 478–480.

Campbell, B. A., and Jaynes, J. 1969. Effect of duration of reinstatement on retention of a visual discrimination learned in infancy. *Developmental Psychology* 1:71–74.

Campbell, B. A., and Randall, P. 1976. Reinstatement of discrimination learning in immature rats as a function of the task component used as the reinstating agent. *Developmental Psychology* in press.

Campbell, B. A., and Spear, N. E. 1972. Ontogeny of memory. *Psychological Review* 79:215–236.

Crespi, L. P. 1942. Quantitative variation in incentive and performance in the white rat. *American Journal of Psychology* 55:467–517.

DiLollo, F. D., and Beez, V. 1966. Negative-contrast effect as a function of magnitude of reward decrement. *Psychonomic Science* 5:99–100.

Flaherty, C. F., Capobianco, S., and Hamilton, L. W. 1973. Effect of septal lesions on retention of negative contrast. *Physiology and Behavior* 11:625–631.

Flaherty, C. F., Riley, E. D., and Spear, N. E. 1973. Effect of sucrose concentration and goal units on runway behavior in the rat. *Learning and Motivation* 4:163–175.

Gatti, S. V., Pais, N., and Weeks, J. R., Jr. 1975. Effects of reinstatement procedures in retention of differential appetitive responding. *Bulletin of the Psychonomic Society* 6(1):57–60.

Gleitman, H., and Steinman, F. 1964. Depression effect as a function of retention interval before and after shift in reward magnitude. *Journal of Comparative and Physiological Psychology* 57:158–160.

Gonzales, R. C., Fernhoff, D., and David, F. G. 1973. Contrast, resistance to extinction, and forgetting in rats. *Journal of Comparative and Physiological Psychology* 84:563–571.

Gordon, W. C., Flaherty, C. F., and Riley, E. P. 1973. Negative contrast as a function of the interval between preshift and postshift training. *Bulletin of the Psychonomic Society* 1:25–27.

Greenfield, H., and Riccio, D. C. 1972. Conditioned reinstatement in rats: Effect of exposure distribution and cue. *Psychological Reports* 31:79–83.

Shubat, E. E., and Whitehouse, J. M. 1968. Reinstatement: An attempt at replication. *Psychonomic Science* 12:215–216.

Silvestri, R., Rohrbaugh, M. J., and Riccio, D. C. 1970. Conditions influencing the retention of learned fear in young rats. *Developmental Psychology* 2: 380–395.

Spear, N. E. 1967. Retention of reinforcer magnitude. *Psychological Review* 74:216–234.

Spear, N. E. 1973. Retrieval of memory in animals. *Psychological Review* 80: 163–194.

Vogel, J. R., Mikulka, P. J., and Spear, N. E. 1968. Effects of shifts in sucrose and saccharin concentration on licking behavior in the rat. *Journal of Comparative and Physiological Psychology* 66:661–666.

AMERICAN JOURNAL OF PSYCHOLOGY
June 1977, Vol. 90, No. 2, pp. 231-242

Similarities of recently acquired and reactivated memories in interference

William C. Gordon
State University of New York at Binghamton

Experiment I demonstrated that when rats were trained on a passive avoidance shortly before training on an active avoidance, subsequent retention of the active avoidance was decreased. Experiment II showed that poor retention also resulted when rats were 'reminded' of the passive avoidance shortly before learning the active avoidance. In both cases, interference with retention was shown to decrease as the interval between passive and active avoidance or between reminder and active avoidance increased. These findings are interpreted as consistent with the notion that newly acquired and reactivated memories share similar characteristics.

Deficits in the retention of a recently learned task can be increased if animals are given prior training in the same apparatus on a task that requires a conflicting response tendency (e.g., Spear, 1971). This 'proactive interference' has been shown to occur when passive-avoidance training precedes active-avoidance training (e.g., Spear, Gordon, and Chiszar, 1972), as well as when that sequence is reversed (e.g., Spear, Gordon, and Martin, 1973). More recently, it has been demonstrated that the degree to which proactive interference occurs depends partially on the interval between the learning of the conflicting tasks, that is, on the intertask interval (Gordon and Spear, 1973a). More specifically, proactive interference was evident after a 1-hr retention interval if prior learning (passive avoidance) preceded later learning (active avoidance) by 30 sec, but not if the intertask interval was 1 or 24 hr. This evidence suggests that for interference to occur, the second task must be learned while the first task is still in 'active' or short-term memory.

However, in contrast to these findings, Gordon and Spear (1973a) also discovered that interference could be produced even with a 24-hr intertask interval if animals were 'reminded' of the passive-avoidance training shortly (i.e., 30 sec) before beginning active-avoidance training. This reminder or 'memory reactivation' treatment consisted of placing each

animal in the training apparatus for a short time without an opportunity to make a response or receive a shock (see Spear, 1973, for a review of studies using the reactivation technique). The purpose of this treatment was simply to 'bias' the animals to retrieve the memory of the passive avoidance without giving them the opportunity for additional learning of the passive avoidance.

Such findings indicate that recently acquired and reactivated memories have similarly interfering effects on the retention of later learning, and they are consistent with other evidence suggesting that recently acquired and reactivated memories have similar characteristics. For example, recently acquired memories are susceptible to the disrupting effects of treatments such as electroconvulsive shock as well as to the facilitating influence of drugs such as strychnine sulphate (see McGaugh and Dawson, 1971). While this susceptibility normally lasts only a matter of minutes after learning, recent studies have shown that both electroconvulsive shock (e.g., Misanin, Miller, and Lewis, 1968) and strychnine (e.g., Gordon and Spear, 1973b) can be effective when administered days after learning, provided they are administered within seconds after the reactivation treatments.

The apparent similarities between recently acquired and reactivated memories suggest that once a memory is reactivated or retrieved, it returns to the same active or short-term state that it occupied shortly after learning (see Lewis, 1969; Gordon and Spear, 1973b). While this interpretation is parsimonious, it has yet to be demonstrated that the characteristics of a reactivated memory dissipate rapidly after reactivation in the same manner that the characteristics of a recently acquired memory dissipate (see McGaugh and Dawson, 1971; Gordon and Spear, 1973a). The fact that the characteristics associated with a recently acquired memory tend to dissipate within a short time has been used to support the notion that time-dependent processes are initiated at the time of learning and are completed shortly after learning. If reactivation or retrieval causes a memory to resume an active or short-term state, it would be reasonable to assume that reactivated memories would be processed in much the same way as newly acquired memories. Thus, it would be predicted that the characteristics of a reactivated memory (such as the ability to interfere with retention of subsequent learning) should dissipate rapidly after a reactivation treatment.

The purpose of the present studies was to test this prediction. Experiment I represents an attempt to determine how long a recently acquired memory remains capable of interfering with retention of subsequent learning. Experiment II explores the same question as it involves a reactivated

memory. If proactive interference is found to dissipate rapidly as the interval between either prior learning or reactivation and later learning increases, these findings will support the notion that reactivation reinitiates short-term or active memory processing.

EXPERIMENT I

METHOD

Subjects and apparatus

The subjects were 35 male albino rats of the Sprague-Dawley strain purchased from Blue Spruce Farms. All animals were 60–100 days old at the time of original training. The primary apparatus, described in detail by Spear et al. (1972), was a one-way avoidance apparatus with a black chamber and a translucent white one. The two chambers, each 27 by 13 by 13 cm, were separated by a door that could be lowered (opened) leaving a 3-cm hurdle between the chambers and simultaneously activating a 2-Hz flashing light (7.5-W) located behind the white chamber. Breaking a photobeam in the middle of the black chamber terminated the light. A scramble footshock provided by a Grason-Stadler shock generator (model E1064GS) could be delivered through the grid floor of either chamber depending on whether active- or passive-avoidance contingencies were in effect.

In addition to the avoidance apparatus, a clear Plexiglas footshock chamber (36 by 36 by 69 cm) was used. The chamber contained a grid floor made of .6-cm diameter rods set 1.7 cm apart. Footshocks could be delivered through this grid floor by the same source used for the avoidance apparatus. Wire-mesh holding cages were used to house animals during both intertrial and intertask intervals.

Design and procedure

Seven rats were randomly assigned to each of five experimental conditions. Animals in four of these conditions underwent the following sequence: passive-avoidance training, active-avoidance training, and finally, a retention test 1 hr later. These four conditions (groups .5P, 5P, 30P, and 60P) differed in terms of the interval that separated passive- and active-avoidance training: .5, 5, 30, and 60 min respectively. Rats under the fifth condition (group .5FS) received the same treatment as animals in group .5P except that a noncontingent footshock treatment replaced original passive-avoidance training.

The procedure for *passive-avoidance* training consisted of placing a rat in the white chamber of the avoidance apparatus so that it faced the chamber door and then opening the door after 3 sec. If the rat crossed over into the black chamber within 60 sec, the door was closed and the rat immediately received a 1.6-mA footshock for 3 sec. If an animal remained in the white chamber for 60 sec, this constituted a successful avoidance and the trial was terminated. Following either a successful avoidance or a footshock, rats were placed in a holding cage for a 30-sec intertrial interval. The criterion for acquisition was two consecutive 60-sec suppressions plus five additional trials. The

footshock treatment given animals in group .5FS consisted of giving the animals noncontingent footshocks in the footshock chamber. These footshocks were of identical intensity and duration as those given during passive-avoidance training in the avoidance apparatus. Also, each animal in group .5FS received the same amount of exposure to the footshock chamber as a member of group .5P received in the avoidance apparatus. The member of group .5P with which each control animal was paired was randomly chosen from the total number of .5P animals run at the time the control animal was to be run. Furthermore, the number and sequence of shocks given each rat in the footshock chamber were the same as received by the paired .5P animal during passive-avoidance training.

The second part of the training sequence, *active-avoidance* training, was identical to the passive-avoidance training except that rats were required to cross from the white to the black compartment within 5 sec after the chamber door opened. Failure to cross within 5 sec resulted in a 1.6-mA footshock that continued until the rat escaped to the black chamber. The criterion for training was three consecutive successful avoidances within 27 trials.

The *retention test* consisted of five trials in which the animals were placed in the white compartment of the avoidance apparatus; after 3 sec, the door was lowered. On these trials, the lowering of the door initiated the flashing light as on the passive- and active-avoidance trials; however, no shocks were administered on any of the trials. Each animal's latency to move to the black chamber was recorded, and a 30-sec intertrial interval was employed. In all cases, animals were housed in a holding cage in the experimental room during the interval between prior learning (passive avoidance or footshock) and subsequent learning (active avoidance). The animals were all returned to their home cages during the 1-hr retention interval.

RESULTS AND DISCUSSION

The median number of trials required to learn passive avoidance was 2 trials for the group 5P and 1 trial for the other three groups. Mann-Whitney U tests performed on these data revealed no significant differences among the four groups that underwent passive-avoidance training [$p > .05$ in all cases]. This result indicates that the four groups were equated as to degree of learning at the end of passive-avoidance training.

The mean number of trials to achieve the active-avoidance criterion for the five treatments was 5.29 for group .5FS, 3.29 for group .5P, 4.14 for group 5P, 8.14 for group 30P, and 3.29 for group 60P. An analysis of variance on these data, excluding those for group .5FS, indicated a significant difference among the four groups [$F(3, 24) = 7.22, p < .01$]. A critical-difference test revealed that group 30P required significantly more trials to learn active avoidance than did any of the other three groups [$p < .01$ in all cases]. Multiple t tests revealed that group 30P also needed significantly more trials to reach criterion than group .5FS [$t(12) = 2.10, p < .05$]. No other groups differed from any other [$p > .05$]. These re-

sults indicate that all groups were not equal in degree of active-avoidance learning prior to the retention test and that the different intertask intervals were responsible for this difference. In any event, the direction of the particular differences obtained for active-avoidance acquisition does not prohibit the conclusions reached below on the retention scores.

As an index of interference with the tendency to actively avoid during the retention test, the crossover latencies for all *five* retention trials were pooled to yield a single mean crossover latency for each animal. Figure 1 represents the mean of these mean crossover latencies for each of the five groups, the longer latencies reflecting greater interference. An analysis of variance was performed on these data, excluding the scores of group .5FS. The analysis revealed a significant difference between the remaining four groups [$F(3, 24) = 3.22, p < .05$]. A critical-difference test showed that groups .5P and 5P exhibited significantly longer latencies than did group 60P [$p < .05$ in both cases]. Multiple t tests were used to compare group .5FS to the other four groups. These tests revealed that group .5FS had significantly shorter latencies than group .5P [$t(12) = 5.40, p < .01$], group 5P [$t(12) = 5.44, p < .05$], and group 30P [$t(12) = 2.85, p < .05$]. There was no significant difference between groups .5FS and 60P

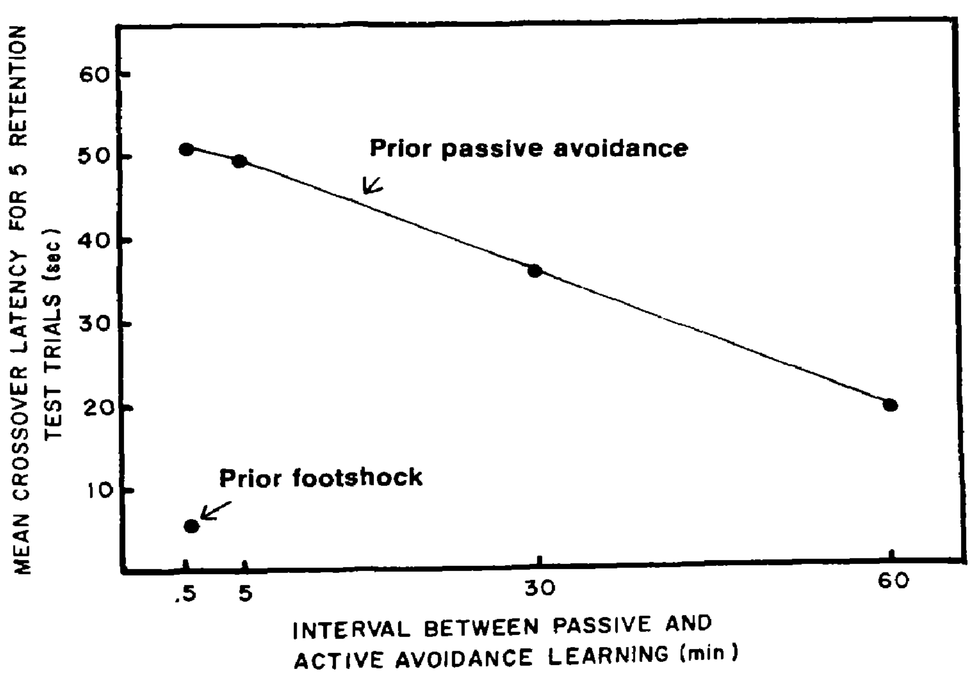


Figure 1. Mean crossover latencies on the five retention trials as a function of the interval between passive- and active-avoidance learning; Experiment I

in terms of the pooled crossover latencies. These results indicate tha the interval between passive-avoidance learning and active-avoida learning increased, there was a progressive decrease in the degree to wl passive-avoidance training interfered with retention of the active-av ance response.

A second measure of the interference with retention of active av ance was the number of trials during the retention test in which anir remained in the white compartment for 60 sec, thus reflecting a tende to perform in accordance with passive- rather than active-avoidance tr ing. The median number of 60-sec suppressions was 0 for group .5F for group .5P, 5 for group 5P, 3 for group 30P, and 0 for group (Mann-Whitney U tests showed that group .5P had significantly more pressions than either group .5FS [$U = 4.5, p < .01$] or group 60P [U 8.5, $p < .05$]. Group 5P had significantly more suppressions than gr .5FS [$U = 4.5, p < .01$] and approached having significantly more pressions than group 60P [$U = 9.5, .10 > p > .05$]. While there wer other significant differences between groups [$p > .05$], group 30P proached having significantly more suppressions than group .5FS [U 9.0, $p = .054$].

A third measure of interference involved the mean crossover later of the groups on the *first* retention trial. Again, longer latencies w indicate a greater tendency to remain passive in accordance with passive-avoidance training. Due to large differences in the indivi group variances, these latencies were subjected to a square root trans mation, and an analysis of variance was performed on the data, exclu those for group .5FS. This analysis revealed a significant difference am the four remaining groups [$F(3, 24) = 3.03, p < .05$]. Critical-differe tests showed that groups .5P and 5P both had significantly longer later than group 60P [$p < .05$]. No other group differences approached nificance. A comparison of group .5FS to the remaining groups by me of t tests showed that group .5FS had significantly shorter latencies t group .5P [$t(12) = 6.65, p < .01$], group 5P [$t(12) = 7.46, p < .01$] group 30P [$t(12) = 2.38, p < .05$]. Groups 60P and .5FS did not d significantly on this measure [$t(12) = 1.76, p > .05$]. As with prev measures, these data indicate a decreasing tendency for passive-avoida behavior on the retention test as the intertask interval increased.

Taken together, the measures of latency and suppression indicate the degree of interference depended on the interval between the pass and active-avoidance learning. According to all the measures asses significant amounts of interference resulted when the intertask inte was 30 min or less. In no case was the amount of interference resul

from an intertask interval of 60 min significant. Supporting this result i the finding from all three reported measures that significantly more inter ference with active-avoidance retention (i.e., higher latencies and mor suppression) occurred with intertask intervals of .5 and 5 min than o 60 min. These data replicate and extend the finding of Gordon and Spea (1973a, Experiment 1) that, again with a retention interval of 1 hr, th degree of memory interference was inversely related to the length of th intertask interval.

EXPERIMENT II

METHOD

Subjects and apparatus

The subjects were 49 rats with characteristics identical to those used Experiment I. The avoidance apparatus, footshock chamber, and holding cag used in Experiment I were also used in Experiment II.

Design and procedure

Seven rats were randomly assigned to each of seven experimental conditior In the four main treatment groups, rats were initially given passive-avoidan training, active-avoidance training 24 hr later, and finally, a retention test 1 later. The procedures for passive- and active-avoidance training and the r tention test were the same as in Experiment I. In addition, animals in tl four main treatment groups were given a treatment intended to reactivate tl memory of passive-avoidance training at some time before the active-avoidan training. For groups .5R, 5R, 30R, and 60R, the interval between reactivati and active-avoidance training was .5, 5, 30, or 60 min respectively. In all cas the reactivation treatment consisted of confining the animals in the black (previously shocked) compartment of the avoidance apparatus for 15 se During this time the chamber door was not opened and no shock was a ministered. While this procedure qualifies as an extinction procedure for passi avoidance, previous studies have shown that confining animals in the previous shocked compartment of the avoidance apparatus for short durations acts reactivate the memory of passive-avoidance learning (Gordon and Spear, 1973a The interval between reactivation and active-avoidance training was spent the second holding cage in the apparatus room. The interval between passiv avoidance training and reactivation, as well as the retention interval, was spe in the home cage.

Aside from these basic treatment groups, three control groups were run. Tl first, group NR, received the same treatment as the four main groups exce that no reactivation treatment was given. These animals spent .5 min in tl second holding cage just before beginning active-avoidance training.

Animals in the other two control groups, groups RP and RFS, were giv noncontingent footshocks in the footshock chamber instead of passive-avoidan training. This treatment was identical to that used in Experiment I except th animals in groups RP and RFS were paired with animals in group .5R. Bo groups received active-avoidance training 24 hr after the footshock procedu

followed 1 hr later by a retention test, and both received a reactivation treatment .5 min before active-avoidance training; however, the groups differed as to the type of reactivation treatment experienced. Group RP was given the same reactivation treatment given the four main treatment groups. This group was included to control for the unconditioned effects of the reactivation treatment. Group RFS received a treatment intended to reactivate the memory of the footshock experience. This involved placing the animals in the footshock chamber for 15 sec with no shock administered. The purpose of this group was to control for the possibility that the reactivation procedure used in the other groups might simply arouse fear in the animals just before active-avoidance acquisition, rather than reactivate the memory of a specific response tendency.

RESULTS AND DISCUSSION

Of the five groups given passive-avoidance training, the median number of trials to learn passive avoidance was 1 trial for groups .5R and 30R and 2 trials for the remaining groups. Mann-Whitney U tests showed that there were no significant differences among these groups on this measure [$p > .05$]. This result indicates that before reactivation all five groups were equal in degree of passive-avoidance learning.

The mean number of trials required to learn active avoidance was 6.43 for group .5R, 7.00 for group 5R, 7.71 for group 30R, 6.86 for group 60R, 8.14 for group RP, and 4.57 for group RFS. An analysis of variance on the mean number of trials required by the four main treatment groups to learn active avoidance revealed no significant differences [$F(3, 24) = .12, p > .05$]. Each of the three control groups was compared to each other and to all other groups by means of t tests. These tests showed that group RP took significantly more trials to learn active avoidance than group RFS [$t(12) = 2.36, p < .05$]; no other groups differed significantly [$p > .05$]. These results indicate that, for the most part, the treatment groups were equated in terms of degree of active-avoidance learning. The one significant difference that did result does not alter the conclusions reached below on the retention scores.

As an index of interference, or a tendency to act in accordance with passive-avoidance training during the retention test, the crossover latencies for all *five* retention trials were pooled to yield a single mean crossover latency for each animal. Figure 2 presents the means of these mean latencies for the seven treatment groups, the longer latencies reflecting greater memory interference. An analysis of variance comparing the four main treatment groups (groups .5R, 5R, 30R, and 60R) on this measure revealed a significant difference among the groups [$F(3, 24) = 9.20, p < .001$]. A comparison of the four individual group means indicated that group .5R had significantly longer crossover latencies than groups 5R,

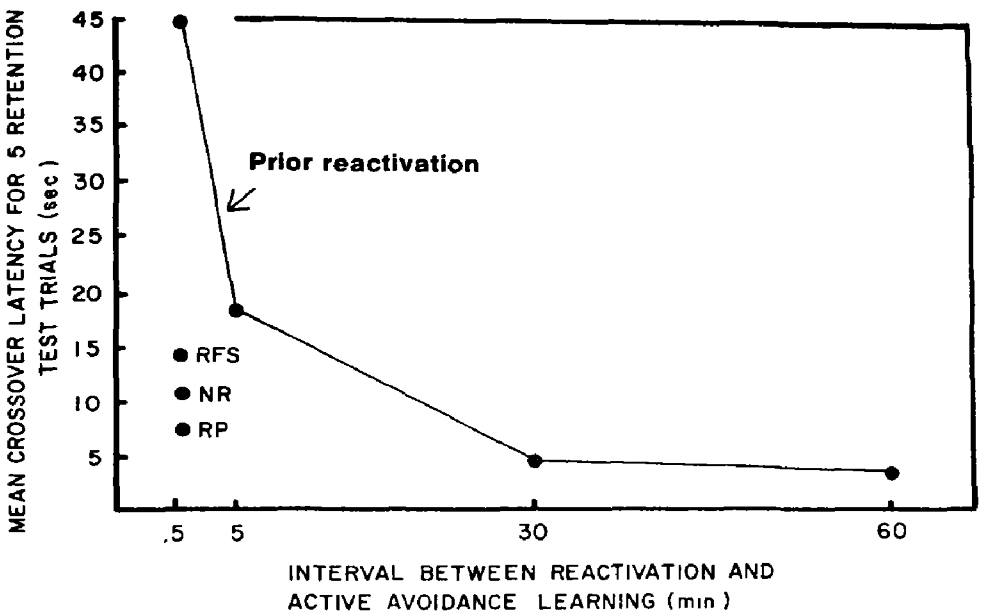


Figure 2. Mean crossover latencies on the five retention trials as a function of the interval between reactivation and active-avoidance learning; Experiment II

30R, and 60R [$p < .01$ in all cases]. There were no other differences among the four groups [$p > .05$]. Each of the control groups (groups NR, RP, and RFS) was then compared to all other groups on this pooled latency measure by means of t tests. These tests showed that group .5R had significantly longer latencies than group NR [$t(12) = 3.05, p < .01$], group RP [$t(12) = 4.28, p < .01$], and group RFS [$t(12) = 2.84, p < .01$]. No other groups differed significantly from the three controls, which did not differ from each other [$p > .05$].

These results indicate that a treatment intended to reactivate the memory of passive-avoidance training was capable of producing interference with retention of a subsequently acquired response only if reactivation occurred just prior (.5 min) to acquisition of the subsequent response. If 5 min or more elapsed between reactivation and active-avoidance acquisition, the interference produced by the reactivation treatment was no greater than that produced by passive-avoidance training given 24 hr before active-avoidance acquisition (groups 5R and NR).

A second measure of interference with retention of active avoidance was the number of trials during the retention test in which animals remained in the white compartment for 60 sec. The median number of sup-

followed 1 hr later by a retention test, and both received a reactivation treatment .5 min before active-avoidance training; however, the groups differed as to the type of reactivation treatment experienced. Group RP was given the same reactivation treatment given the four main treatment groups. This group was included to control for the unconditioned effects of the reactivation treatment. Group RFS received a treatment intended to reactivate the memory of the footshock experience. This involved placing the animals in the footshock chamber for 15 sec with no shock administered. The purpose of this group was to control for the possibility that the reactivation procedure used in the other groups might simply arouse fear in the animals just before active-avoidance acquisition, rather than reactivate the memory of a specific response tendency.

RESULTS AND DISCUSSION

Of the five groups given passive-avoidance training, the median number of trials to learn passive avoidance was 1 trial for groups .5R and 30R and 2 trials for the remaining groups. Mann-Whitney U tests showed that there were no significant differences among these groups on this measure [$p > .05$]. This result indicates that before reactivation all five groups were equal in degree of passive-avoidance learning.

The mean number of trials required to learn active avoidance was 6.43 for group .5R, 7.00 for group 5R, 7.71 for group 30R, 6.86 for group 60R, 8.14 for group RP, and 4.57 for group RFS. An analysis of variance on the mean number of trials required by the four main treatment groups to learn active avoidance revealed no significant differences [$F(3, 24) = .12$, $p > .05$]. Each of the three control groups was compared to each other and to all other groups by means of t tests. These tests showed that group RP took significantly more trials to learn active avoidance than group RFS [$t(12) = 2.36$, $p < .05$]; no other groups differed significantly [$p > .05$]. These results indicate that, for the most part, the treatment groups were equated in terms of degree of active-avoidance learning. The one significant difference that did result does not alter the conclusions reached below on the retention scores.

As an index of interference, or a tendency to act in accordance with passive-avoidance training during the retention test, the crossover latencies for all *five* retention trials were pooled to yield a single mean crossover latency for each animal. Figure 2 presents the means of these mean latencies for the seven treatment groups, the longer latencies reflecting greater memory interference. An analysis of variance comparing the four main treatment groups (groups .5R, 5R, 30R, and 60R) on this measure revealed a significant difference among the groups [$F(3, 24) = 9.20$, $p < .001$]. A comparison of the four individual group means indicated that group .5R had significantly longer crossover latencies than groups 5R,

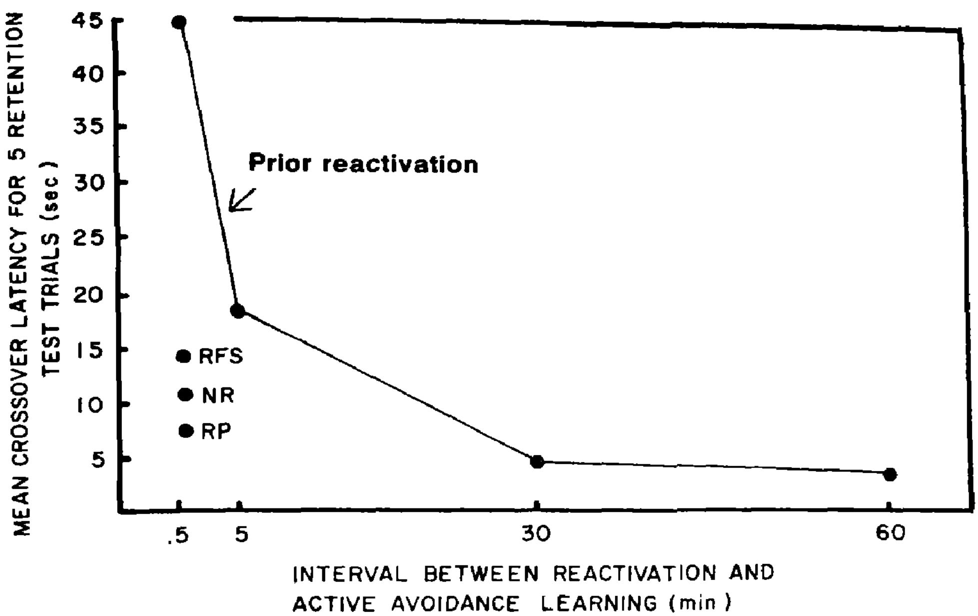


Figure 2. Mean crossover latencies on the five retention trials as a function of the interval between reactivation and active-avoidance learning; Experiment II

30R, and 60R [$p < .01$ in all cases]. There were no other differences among the four groups [$p > .05$]. Each of the control groups (groups NR, RP, and RFS) was then compared to all other groups on this pooled latency measure by means of t tests. These tests showed that group .5R had significantly longer latencies than group NR [$t(12) = 3.05, p < .01$], group RP [$t(12) = 4.28$, $p < .01$], and group RFS [$t(12) = 2.84$, $p < .01$]. No other groups differed significantly from the three controls, which did not differ from each other [$p > .05$].

These results indicate that a treatment intended to reactivate the memory of passive-avoidance training was capable of producing interference with retention of a subsequently acquired response only if reactivation occurred just prior (.5 min) to acquisition of the subsequent response. If 5 min or more elapsed between reactivation and active-avoidance acquisition, the interference produced by the reactivation treatment was no greater than that produced by passive-avoidance training given 24 hr before active-avoidance acquisition (groups 5R and NR).

A second measure of interference with retention of active avoidance was the number of trials during the retention test in which animals remained in the white compartment for 60 sec. The median number of sup-

pressions made by group .5R was 4, while all other groups had a median of 0. Mann-Whitney *U* tests revealed that group .5R had significantly more suppressions than group 30R [$U = 4, p < .01$], group 60R [$U = 3.5, p < .01$], group RP [$U = 4.5, p < .01$], and group RFS [$U = 7, p < .05$]. Group .5R also approached having significantly more suppressions than group 5R [$U = 10.5, .09 > p > .05$] and group NR [$U = 9, p = .054$]. No other group differences approached significance [$p > .05$]. These results support the conclusion that significant interference with retention of active-avoidance learning occurred only when the reactivation treatment was presented shortly (.5 min) before active-avoidance training.

Unlike the results on pooled latencies and number of suppressions, however, an analysis of variance of the latencies on the *first* retention trial revealed no significant differences among the four main treatment groups [$F(3, 24) = .55, p > .05$]. Likewise, *t* tests showed that none of the main treatment groups differed significantly from any of the three control groups [$p > .05$]. Thus, there were no significant differences among any of the treatment groups in terms of latencies on the first retention trial. It should be noted, however, that the pattern of results on this measure was very similar to that for the pooled latencies. The mean retention latencies for the seven treatment groups (in sec) were 23.61 for group .5R, 15.46 for group 5R, 12.10 for group 30R, 10.07 for group 60R, 12.30 for group NR, 10.85 for group RP, and 12.06 for group RFS. Thus, this measure, while not statistically supporting the first two measures, does suggest a similar pattern of effects.

Therefore, several conclusions may be drawn from the results of this experiment. First, using a 24-hr interval between passive- and active-avoidance learning, retention of active avoidance after 1 hr was significantly impaired if the memory of passive avoidance was reactivated shortly (.5 min) before active-avoidance acquisition. Little interference was evidenced if no reactivation treatment was given (group NR). When the interval between reactivation and active avoidance learning was extended (to 5, 30, or 60 min), a significant reduction in the amount of interference resulted. The pattern of results for all three measures of interference is that as the interval between reactivation and active-avoidance learning increased, interference correspondingly decreased. The reactivation treatment itself was shown to have little influence on active-avoidance retention (group RP) unless passive-avoidance learning had preceded it. Furthermore, an attempt to reactivate the memory of a prior noncontingent footshock experience (group RFS) also failed to produce interference with retention of active-avoidance learning.

GENERAL DISCUSSION

Together, the present studies replicate and extend Gordon and Spear's (1973a) findings that proactive interference decreases as the interval between prior and subsequent learning increases and that reactivation of a prior memory just before subsequent learning significantly increases the proactive interference due to the prior learning. More important, however, is the further finding that proactive interference due to either prior learning or reactivation decreases rapidly as the intertask interval or the interval between reactivation and subsequent learning increases. This finding suggests that while newly acquired and reactivated memories are capable of producing interference with retention, this capacity diminishes rapidly as the interval following original learning or reactivation increases.

These data are consistent with a growing body of evidence that newly acquired and reactivated memories share similar characteristics (e.g., Misanin et al., 1968; Gordon and Spear, 1973a, 1973b; Gordon, 1973). One interpretation of such similarities is that once a memory is reactivated, it returns to an active or short-term state and is reprocessed in much the same manner as a newly acquired or 'short-term' memory. This type of model, first suggested by Lewis (1969), attributes to the mechanisms of memory a dynamic character not present in a simple consolidation model (McGaugh, 1966). According to the consolidation model, once a memory has achieved permanent storage, it is in a passive state and relatively immune to modification by a source short of massive insult to the brain. But if reactivation causes a passive memory to return to an active state, there is a basis for understanding how organisms are able to continually add to, delete from, or otherwise modify memories that have been stored long ago.

Notes

This research forms part of a dissertation submitted to Rutgers — The State University in partial fulfillment of the requirements for the Ph.D. The author acknowledges the guidance of Norman E. Spear, who directed the dissertation, and the helpful comments of M. R. D'Amato, R. R. Miller, and L. H. Hamilton. This research was supported by Grants MH-18619 and MH-47359 from the National Institute of Mental Health to Norman E. Spear and by Predoctoral Fellowship Grant MH-54528 from the National Institute of Mental Health to the author. Requests for offprints should be sent to William C. Gordon, Department of Psychology, State University of New York at Binghamton, Binghamton, New York 13901. Received for publication November 18, 1975; revision, June 7, 1976.

References

Gordon, W. C. 1973. Similarities between short-term and reactivated memories. Doctoral dissertation, Rutgers — The State University.

Gordon, W. C., and Spear, N. E. 1973a. The effect of reactivation of a previously acquired memory on the interaction between memories in the rat. *Journal of Experimental Psychology* 99:349–355.

Gordon, W. C., and Spear, N. E. 1973b. The effects of strychnine on recently acquired and reactivated passive avoidance memories. *Physiology and Behavior* 10:1071–1075.

Lewis, D. J. 1969. Sources of experimental amnesia. *Psychological Review* 76: 461–472.

McGaugh, J. L. 1966. Time dependent processes in memory storage. *Science* 153:1351–1358.

McGaugh, J. L., and Dawson, R. G. 1971. Modification of memory storage processes. In *Animal memory*, ed. W. K. Honig and H. James. New York: Academic Press.

Misanin, J. R., Miller, R. R., and Lewis, D. J. 1968. Retrograde amnesia produced by electroconvulsive shock after reactivation of a consolidated memory trace. *Science* 160:554–555.

Spear, N. E. 1971. Forgetting as retrieval failure. In *Animal memory*, ed. W. K. Honig and H. James. New York: Academic Press.

Spear, N. E. 1973. Retrieval of memory in animals. *Psychological Review* 80: 163–194.

Spear, N. E., Gordon, W. C., and Chiszar, D. A. 1972. Interaction between memories in the rat: Effect of degree of prior conflicting learning on forgetting after short intervals. *Journal of Comparative and Physiological Psychology* 78:471–477.

Spear, N. E., Gordon, W. C., and Martin, P. A. 1973. Warm-up decrement as failure in memory retrieval in the rat. *Journal of Comparative and Physiological Psychology* 85:601–614.

AMERICAN JOURNAL OF PSYCHOLOGY
June 1977, Vol. 90, No. 2, pp. 243–251

The use of relief and nonrelief from shock in the double alleyway

Joseph A. Sgro
Virginia Polytechnic Institute and State University

Performance in a double alleyway was investigated using relief from shock instead of a food reward. Rats were given shock with either long (.90 sec) on periods and short (.10 sec) off periods or vice versa. During 25 training trials, entry into goalbox 1 terminated shock in alley 1. After 8 sec of relief, shock was delivered in goalbox 1 and remained on until escape into goalbox 2. During 18 test trials, a 50% relief schedule was given in goalbox 1. The typical frustration effect — faster alley 2 speeds after nonrelief — was found for the rats with the short on period. A reversed frustration effect — slower alley 2 speeds after nonrelief — was found for the rats with a long on period. Implications for frustration theory are discussed.

The motivational properties of frustrative nonreward have typically been demonstrated in alley 2 of a double alleyway. With 50% partial reinforcement in goalbox 1, rats run faster after nonreward than after reward (Amsel, 1958; Wagner, 1959). The finding that alley 2 performance is enhanced after nonreward has been designated the *frustration effect*.

Recently, three studies (Graham and Longstreth, 1970; Lambert and Hammond, 1970; Millard and Woods, 1975) have attempted to extend the concept that relief and nonrelief from an aversive stimulus in alley 1 of a double alleyway constitute analogous conditions of reward and nonreward respectively. Of these studies, the generality of the frustration effect with food reward to relief from an aversive stimulus was confirmed by Millard and Woods, who found faster swimming speeds after nonrelief than after relief from cold water in goalbox 1 of a double waterway. A second confirmation came from Graham and Longstreth, who found that with distinctive goalbox 1 cues signaling the presence of shock and no shock, a shift to shock for the cues previously associated with no shock produced faster alley 2 speeds. Contrary to the first two studies, Lambert and Hammond found that after a series of trials with continuous relief from shock in goalbox 1, a shift to 50% relief resulted in slower alley 2

speeds after relief from shock, a finding which was categorized as a *reversed* frustration effect.

The present study focused on the concern that the discrepant finding by Lambert and Hammond could have been peculiar to the use of continuous shock as the aversive stimulus. Two considerations are of particular importance to this concern. First, studies in avoidance learning (D'Amato, Keller, and Biederman, 1965; D'Amato, Keller, and DiCara, 1964) have revealed that the use of continuous shock has a disruptive effect on behavior. Second, in a pilot study replicating Lambert and Hammond's procedure, the present author found that the majority of rats showed extreme freezing behavior in goalbox 1 on nonrelief trials and tenacious biting behavior upon release from goalbox 2. Thus, the present study was designed to determine whether Lambert and Hammond's finding was peculiar to the use of severe continuous shock in the double alleyway. Discontinuous shock of long (.90 sec) on periods and short (.10 sec) off periods was used instead of continuous shock. In addition, a second group received short (.10 sec) on periods and long (.90 sec) off periods. Both groups received 25 trials with relief from shock in goalbox 1 and then were shifted to 18 trials with relief from shock in goalbox 1 on a random 50% of those trials.

METHOD

Subjects

The subjects were 28 male hooded rats obtained from the laboratory colony of the Department of Psychology at Virginia Polytechnic Institute and State University. The subjects were 90–100 days old at the beginning of the experiment. Food and water were freely available in the home cage of each rat.

Apparatus

The apparatus was a double alleyway consisting of two straight alleyways in series; it was constructed of 1.91-cm plywood and was 6.45 cm wide and 10.16 cm high throughout. The first startbox and the first alley were 30.48 cm and 185.41 cm long respectively. The first goalbox, which also served as the second startbox, was 36.38 cm long. The second alley was 135.89 cm long. The second goalbox, which was L-shaped, was 20.86 cm long, and the arm of the L was 30.48 cm long. Guillotine doors separated the startbox, runway, and goalbox sections of each alleyway. The entire apparatus was painted flat black and was covered by .32-cm Plexiglas. Lighting was provided by 7-W bulbs suspended 96.52 cm above the apparatus.

The surface for shocking the rats consisted of two L-shaped aluminum runners, each of which formed one wall and half of the floor. A 1.58-cm gap ran the entire length of the apparatus and separated the floor portion of the two runners. A 1,200-V ac transformer and a large series resistance capable of being adjusted to values of 2.18, 1.85, 1.60, and 1.41 MΩ provided currents of

.55, .65, .75, and .85 mA respectively, as measured by a milliammeter across the runners. The shock's duration was controlled by a timer that allowed alternating durations of either .90 sec on and .10 sec off or .10 sec on and .90 sec off.

Start times for alley 1 were measured from the opening of the startbox door to the interruption of a photobeam 25.40 cm distant. *Running times* for alley 1 were measured from the breaking of the start-time beam to the interruption of a second beam 88.90 cm distant. Start times for alley 2 were measured from the opening of the door leading into alley 2 until the breaking of a beam 25.40 cm distant. Running times for alley 2 were measured from the breaking of the latter beam to the interruption of a second beam 120.65 cm distant.

Procedure

The subjects were randomly assigned to two preshift groups: group L ($N = 14$) received shocks that were long on and short off; group S ($N = 14$) received shocks that were short on and long off. Both groups received relief from shock in goalbox 1 for 25 trials. All subjects then received 18 test trials under random 50% relief in goalbox 1.

During the first phase, the typical procedure involved placing a rat in the first startbox and raising the door when the rat was oriented to it. When the door was raised, shock was simultaneously applied to the startbox and alley 1. When the alley 1 photobeam that measured alley 1 run times and was located 7.62 cm within goalbox 1 was interrupted, the rat was detained in goalbox 1 for 8 sec without shock. Following the 8-sec relief, shock was applied simultaneously to the runners of goalbox 1 and alley 2. After 2 sec, the guillotine door leading to alley 2 was raised and the rat was allowed to enter a safe goalbox 2. Since the timer controlling the shock's durations was operating continually, the point of the shock's onset at the initiation of the period for shock varied randomly within trials for each group. However, since the alternating sequence of durations occurred in a constant cycle, group L always received a total of 1.80 sec of shock and group S received a total of .20 sec of shock.

Beginning with trials 1 through 5, a .55-mA shock was administered and then increased in .10-mA steps for each successive block of five trials. At trial 16, shock was increased to .85 and was maintained at that level for the remainder of the experiment. During the second phase, 50% of the trials were administered under the same procedures for relief used during the first phase. The other 50% of the trials involved nonrelief from shock in goalbox 1. During these latter trials, goalbox 1 was no longer safe and the rat was detained in goalbox 1 for 2 sec with shock. The door leading to alley 2 and goalbox 2 was then opened and the rat was allowed to escape to a safe goalbox 2. The order of relief (R) and nonrelief (N) trials followed the same pattern used by Lambert and Hammond (1970) and by Amsel and Roussel (1952): NRN, RNR, NNR, RRN, NRR, RNN. All trials for each subject were completed in one session, and the intertrial interval was 30–45 sec.

RESULTS

Start and running times for each trial were converted to speeds (1/sec). For alley 1 start and running speeds, median speeds were determined for

each subject over each block of five trials for the training phase and over each block of six trials for the test phases. For alley 2 start and running speeds, median speeds were computed for each subject over each block of five trials during the training phase and over each block of three trials for each experimental condition over the test phase.

Training session

Mean median *alley 1* start and running speeds are presented in Figure 1. It may be seen in Figure 1 that group L showed faster alley 1 start and running speeds over blocks 1 through 3. It is apparent, however, that group L dropped below group S over the remainder of the study. Analyses of variance performed on the mean median alley 1 start and running speeds over blocks 1 through 5 revealed a significant interaction of blocks by groups for start [$F(4, 104) = 4.25$, $p < .05$] and running speeds [$F(4,$

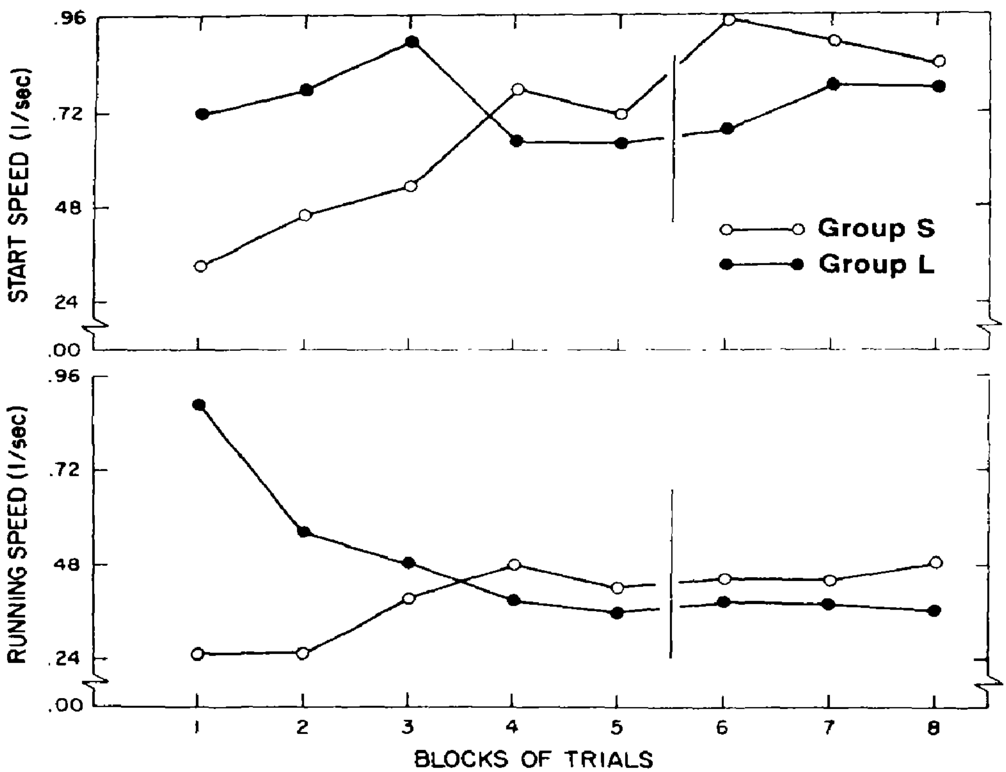


Figure 1. Mean median alley 1 start and running speeds as a function of blocks: for the training session, plotted in blocks of five trials; for the test session, in blocks of six

104) = 19.42, $p < .01$]. Assessment of these interactions revealed that over blocks 1 through 3, group L showed faster start and running speeds than group S [$F(1, 26) = 10.14$ and 10.01, $p < .01$]. Analyses over blocks 4 and 5, however, indicated that before the test session, the groups failed to differ significantly [$Fs < 1$].

Mean median *alley 2* start and running speeds are presented in Figure 2. Observation of the alley 2 start speeds indicates that during training group L maintained faster speeds upon being allowed access to alley 2. It is evident from Figure 2 that for the running speeds, group L was faster over the initial blocks of training but did not differ from group S over the later blocks of trials. Analysis of variance performed on the alley 2 start speeds over trial blocks 1 through 5 indicated that the effect of groups was statistically reliable [$F(1, 26) = 7.36$, $p < .05$], as was that of blocks [$F(4, 104) = 18.37$, $p < .01$], but the interaction of blocks by groups failed to achieve statistical significance [$F(4, 104) = 1.45$, $p > .05$]. For running speeds over blocks 1 through 5, analysis of variance revealed a significant interaction of groups by blocks [$F(4, 104) = 20.53$, $p < .01$]. Group L ran significantly faster than group S over blocks 1 and 2 [$F(1, 26) = 11.16$, $p < .01$], but not over the last three blocks of training [$F < 1$].

Test session

Analyses of variance performed on *alley 1* start and running speeds over blocks 6 through 8 revealed that neither the effect of groups nor that of blocks was significant for either measure [$Fs < 1$]. The interaction of blocks by groups failed to be statistically reliable for either start [$F(2, 52) = 1.38$, $p > .05$] or running speeds [$F < 1$].

It is evident from Figure 2 that after the shift to 50% relief in goalbox 1, group L showed slower *alley 2* start speeds after nonrelief than after relief from discontinuous shock. A most revealing feature of Figure 2, however, is that over the same block of trials, group S, which received shocks of short on durations, showed faster alley 2 start speeds after nonrelief of discontinuous shock in goalbox 1. Analysis of variance performed over blocks 6 through 8 on mean median alley 2 start speeds revealed that only the interaction of relief and nonrelief by duration was statistically significant [$F(1, 26) = 18.24$, $p < .01$]. Simple-effects analyses indicated that group L started significantly slower after nonrelief than after relief of shock in goalbox 1 [$F(1, 26) = 10.63$, $p < .01$]. In addition, it was found that group S started significantly faster after nonrelief than after relief of shock in goalbox 1 [$F(1, 26) = 7.73$, $p < .01$]. Further-

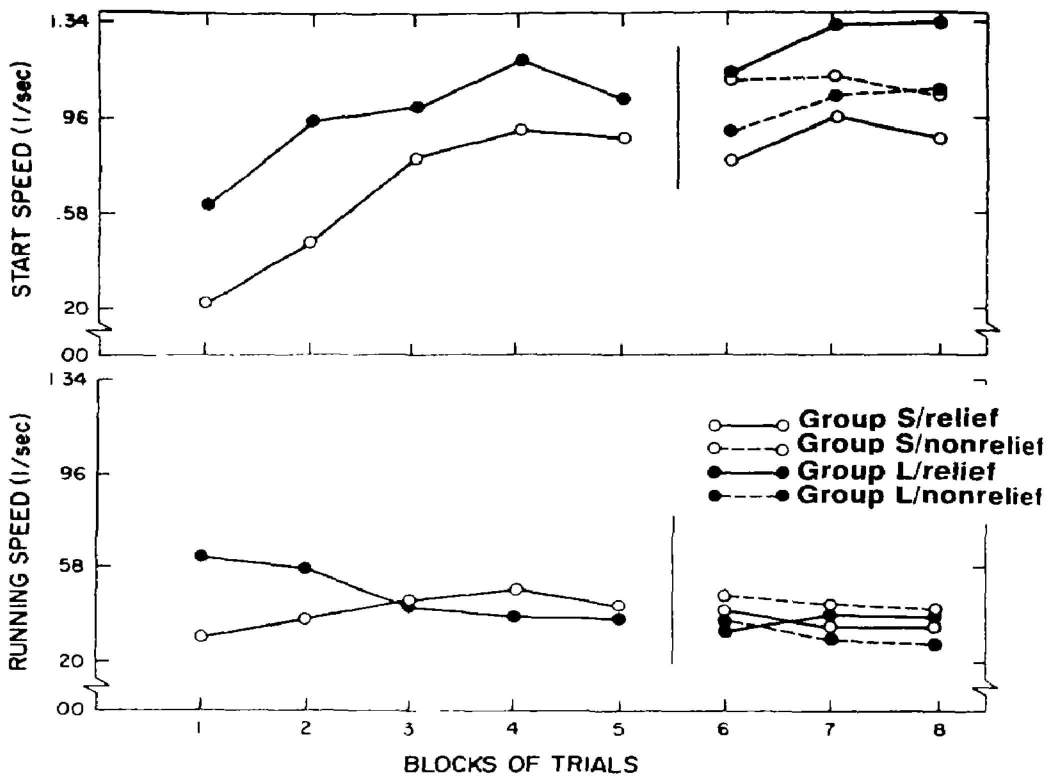


Figure 2. Mean median alley 2 start and running speeds as a function of blocks: for the training session, plotted in blocks of five trials; for the within-subjects conditions during the test session, in blocks of three

more, group L started significantly faster than group S after relief [$F(1, 26) = 6.14$, $p < .05$]. No significant differences were found, however, after nonrelief [$F < 1$].

Inspection of the alley 2 running speeds in Figure 2 shows that while group S ran faster after nonrelief than after relief, the magnitude of the difference appeared to be smaller than for the start speeds. Group L did not show a difference in running speeds until later in the test session. Analysis of variance performed on alley 2 running speeds over blocks 6 through 8 revealed that the only significant findings were an interaction of relief by duration [$F(1, 26) = 9.75$, $p < .01$] and an interaction of relief by blocks [$F(2, 52) = 3.26$, $p < .05$]. Subsequent analyses revealed that group S ran significantly faster in alley 2 after nonrelief than after relief from shock over blocks 6 through 8 [$F(1, 26) = 5.84$, $p < .05$]. Group L, on the other hand, showed no difference in alley 2 running

speeds after relief and nonrelief on block 6 [$F < 1$] or block 7 [$F(1, 78) = 3.74$, $p > .05$]. On block 8, however, group L ran significantly slower after nonrelief than after relief in goalbox 1 [$F(1, 78) = 8.80$, $p < .01$].

DISCUSSION

The primary purpose of the present study was to determine whether the failure of Lambert and Hammond (1970) to obtain faster alley 2 performance after nonrelief than after relief from shock might have been due to the peculiarity of continuous shock. The present data indicate that with discontinuous shock of long on periods (.90 sec, group L), alley 2 start speeds were faster after relief than after nonrelief. These data agree with Lambert and Hammond's. On the other hand, rats that received discontinuous shock of short on periods (.10 sec, group S) showed faster alley 2 start and running speeds after nonrelief than after relief from shock in goalbox 1. These findings agree with Graham and Longstreth's (1970) and with Millard and Woods's (1975) findings of faster alley 2 performance after nonrelief from an aversive stimulus.

The likelihood, however, that the processes underlying a shift from relief to nonrelief of shock are similar to the processes underlying enhanced alley 2 performance when a food reward is completely removed may be questioned. In studies on double-alleyway escape conditioning, the critical variable that distinguishes relief from nonrelief is the interval between the offset of shock in alley 1 and the *onset* of shock in goalbox 1 and alley 2. Thus, the eventual presence of shock (or aversive stimuli) in goalbox 1 on both relief and nonrelief trials represents a distinctly different situation than is encountered when food reward is completely removed in demonstrating the traditional frustration effect (Amsel, 1958, 1962).

Within frustration theory, the most distinctive occasion for the generation of frustration would probably be the denial of the *offset* of shock on nonrelief trials in goalbox 1. As a consequence of the contribution of frustration to the motivational level, alley 2 speeds after nonrelief would be expected to be faster than those after relief. This hypothesis is supported by group S. The reversed frustration effect for group L might be interpreted along the lines offered by McAllister, McAllister, Brooks, and Goldman (1972) to account for the effects of a shift in the reward's magnitude (fear reduction) on hurdle jumping to escape fear-eliciting stimuli. A shift from a large reward to a small reward resulted in a degradation in performance, which was attributed to incompatible responses elicited by primary frustration brought about by the disconfirmation of relief. In

the present study, the interfering effects of frustration would be compounded by the severity of the shocks with the long on periods.

An interesting finding was that group L showed faster alley 2 start speeds than group S after relief from shock. The interval between the shock's offset and its onset would offer the opportunity for anticipatory pain responses (r_p) or fear to be generated in goalbox 1. Considering that the total amount of shock for group L (1.80 sec) was greater than for group S (.20 sec), the faster alley 2 start speeds for group L may easily be attributed to the combination of fear and a direct motivational response to shock.

The alley 2 running speeds also reveal enhanced performance for group S after nonrelief and, later in the test session, depressed performance for group L after nonrelief. The observation that alley 2 running speeds were less sensitive to prior experience with shock in goalbox 1 has also been reported by Graham and Longstreth (1970). Apparently, the closer the animal to a safe goalbox 2, the less influence exerted by shock in goalbox 1.

The data on training merit consideration. As may be seen in Figure 1, group L showed a decrement in alley 1 start and running speeds relative to group S. While similar data has been reported by Graham and Longstreth (1970) and has been attributed to habituation to shock, the superior performance exhibited by group L in alley 2 would negate this possibility. An alternative explanation for the decrement might be that group L experiences a greater approach/avoidance conflict in alley 1.

Notes

The author thanks David R. Miller, for his assistance in constructing the apparatus, and Howard A. Beiles, Nancy Dailey Beiles, and Janet L. Schmid, for their assistance in the collection of data. Offprint requests should be sent to Joseph A. Sgro, Department of Psychology, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24060. Received for publication August 25, 1975, revision, June 11, 1976.

References

Amsel, A. 1958. The role of frustrative nonreward in noncontinuous reward situations. *Psychological Bulletin* 55:102–119.

Amsel, A. 1962. Frustrative nonreward in partial reinforcement and discrimination learning: Some recent history and a theoretical extension. *Psychological Review* 69:306–328.

D'Amato, M. R., Keller, D., and Biederman, G. 1965. Discriminated avoidance learning as a function of parameters of discontinuous shock. *Journal of Experimental Psychology* 70:543–548.

D'Amato, M. R., Keller, D., and DiCara, L. V. 1964. Facilitation of discriminated avoidance learning by discontinuous shock. *Journal of Comparative and Physiological Psychology* 58:344–349.

Graham, T. M., and Longstreth, L. E. 1970. The role of frustrative nonreward in instrumental escape conditioning. *Journal of General Psychology* 83: 185–192.

Lambert, J. V., and Hammond, L. J. 1970. Effect of frustrative nonrelief upon shock-escape behavior in the double runway. *Journal of Experimental Psychology* 85:216–219.

McAllister, D. E., McAllister, W. R., Brooks, C. I., and Goldman, J. A. 1972. Magnitude and shift of reward in instrumental and aversive learning in rats. *Journal of Comparative and Physiological Psychology* 80:490–501.

Millard, W. J., and Woods, P. J. 1975. Frustrative nonrelief in instrumental escape conditioning. *Animal Learning and Behavior* 3:28–32.

Wagner, A. R. 1959. The role of reinforcement and nonreinforcement in an "apparent frustration effect." *Journal of Experimental Psychology* 57:130–136.

AMERICAN JOURNAL OF PSYCHOLOGY
June 1977, Vol. 90, No. 2, pp. 253–262

Organizational determinants of subjective contour: The subjective Necker cube

Drake R. Bradley and Heywood M. Petry
Bates College

With specially arranged inducing elements on a white surface of uniform luminosity, a phenomenally complete Necker cube can be seen in an array where only the 'corners' of the cube are physically represented. The subjectively seen bars of the cube disappear when the inducing 'discs' are seen as 'holes' in an interposing surface, through which the corners of a partially occluded cube are viewed. Illusory brightness effects are also observed in connection with the different organizations of this ambiguous figure.

In this report, we demonstrate a perceptual phenomenon with important implications for theories of subjective contour. A full description of the phenomenon is provided, followed by some preliminary data collected to substantiate its existence in a naive sample. We then conclude with a brief discussion of the theoretical significance of the visual effects observed.

If the reader will invest a few minutes in carefully observing Figure 1, a number of interesting and rather striking visual effects may be seen. First, and most prominent, is the impression of a three-dimensional cube-like object suspended in space such that each 'corner' of the cube is in front of a black 'disc.' The bars connecting the corners of the cube can be seen extending beyond, and therefore between, the discs. They have the appearance of a slightly brightened or more intense area bounded by faint contours or edges the same distance apart as the edges of the bars actually over a disc. Since this display was produced by placing black inducing elements in special arrangements on a white surface of uniform luminosity, the contours seen extending between the discs do not actually exist in the physical array. Consequently, the perception of the cube in its entirety, and in particular those portions seen between the discs, is illusory. Nevertheless, this subjectively seen cube manifests many of the defining properties of a standard Necker cube, such as the strong impression of three-dimensionality, spontaneous reversals in apparent orienta-

tion and depth, and a slight difference in apparent size of the front and back faces of the cube.

The subjective Necker cube (Bradley, Dumais, and Petry, 1976) is a three-dimensional variation of a phenomenon, long known, called subjective contour (Schumann, 1904). A standard configuration for the perception of subjective contour is presented in Figure 2 (Kanizsa, 1955). The 'sides' of a triangle are perceived as faint contours or edges extending from one black disc to the next. As with the subjective Necker cube, the perception of these contours is illusory. Further, the triangle bounded by these subjective contours appears slightly brighter than the background on which the triangle rests, even though the reflectances are equal. Figures 1 and 2 are similar in that they both produce phenomenally complete objects (cube, triangle) that appear brighter than their backgrounds and are bounded by illusory contours between the inducing elements.

However, in one important respect the visual arrays in Figures 1 and 2 are different. In Figure 1, an alternative perceptual organization is possible, whereby the previously noted subjective contours disappear, to be replaced by a new set of illusory contours elsewhere in the display. The

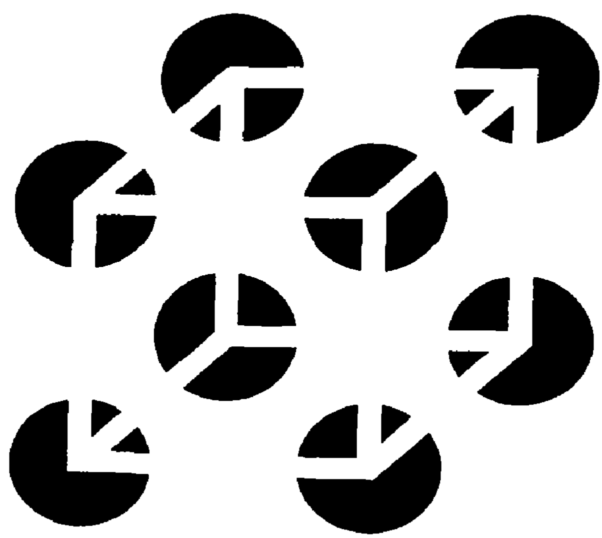


Figure 1. The subjective Necker cube. A phenomenally complete Necker cube can be seen overlying a white surface and eight black discs; so viewed, illusory contours corresponding to the bars of the cube can be seen extending between the discs. The illusory bars of the cube disappear when the discs are seen as 'holes' in an interposing surface, through which the corners of a partially occluded cube are viewed; curved subjective contours are then seen demarcating the interior edges of the 'holes'

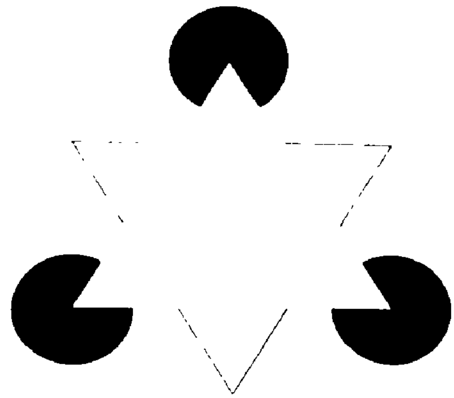


Figure 2. A standard configuration for the perception of subjective contour: the 'sides' of the triangle seen extending between the discs are illusory, as is the difference in apparent brightness of the triangle and the white background

alternative organization is simply this: imagine you are looking into a dark room through eight 'holes' in an interposing white surface. A cube is suspended in space in *back* of the surface such that each corner of the cube is visible through one of the holes but the rest of the cube is occluded by the interposing surface. With this cube-in-back organization, the subjective contours previously seen between the discs are not present: instead, the bars of the cube are now amodally completed behind the occluding white surface.[1] However, careful observation reveals that new subjective contours are present during this cube-in-back organization. In each case where the interior edge of a 'hole' traverses a bar of the cube in back, a curved subjective contour can be seen.

As might be expected, these curved subjective contours disappear when the first, cube-in-front organization prevails. Note that in both cube-in-front and cube-in-back organizations, subjective contours are seen where the interpolation of an edge or boundary by the visual system serves to create a phenomenally complete object consistent with the prevailing organization (i.e., either the 'cube' of the cube in front, or the 'holes' of the occluding surface of the cube in back). These observations suggest that central organizing factors are important in determining the perception of such illusory contours (Bradley and Dumais, 1976).

Figure 3 presents unambiguous versions of the cube-in-front (B) and cube-in-back (C) organizations of the subjective Necker cube, along with a 'straddled' organization (D) that is occasionally reported. Real contours are filled in over those regions of the display where subjective contours would be seen if the ambiguous version were viewed under the same

organization. It should be noted that the unambiguous organizations depicted in Figure 3 misrepresent the appearance of subjective contours insofar as they *imply* that black lines are seen. Rather, the experience of a subjective contour is generally described as that of an edge or border separating a discrete change in apparent brightness — as a one-step change in luminance, rather than the two-step change evident with a black line on a white background.

One additional and very striking effect may be observed in Figure 1. After you have gained some facility in seeing both of the organizations described above, carefully direct your attention to the apparent brightness of the cube while in each of the two positions. When the cube is organized in back of the surface, it appears extremely bright; in fact, it may even appear luminous. When the cube is organized in front of the

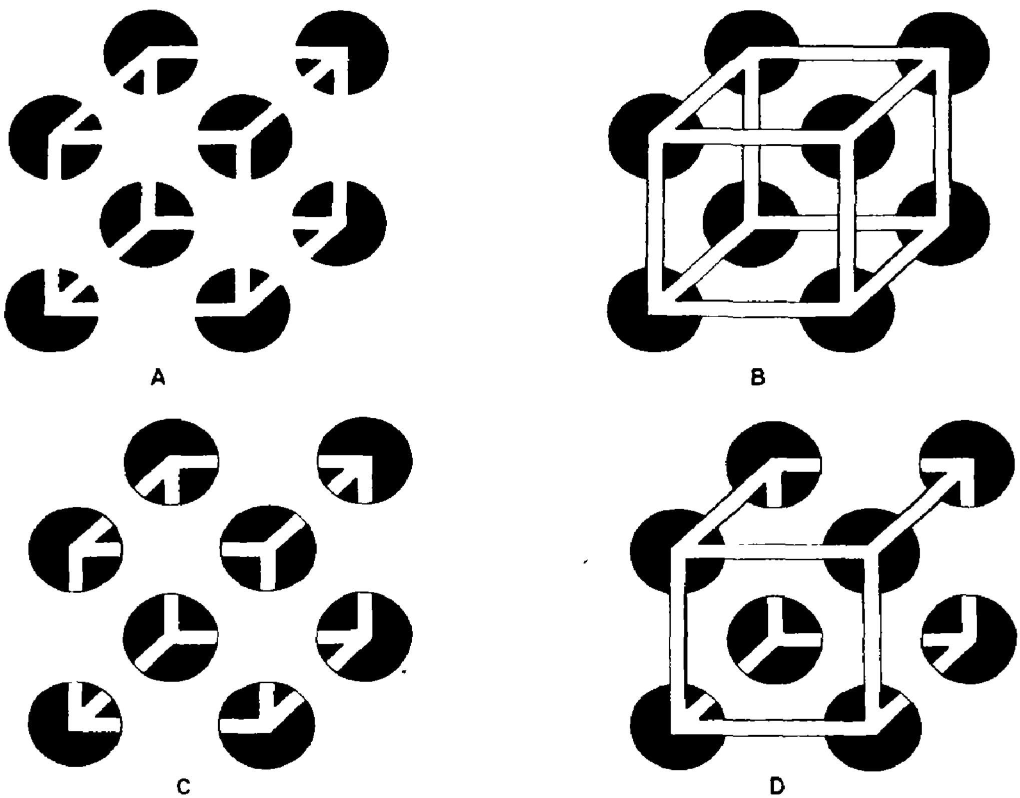


Figure 3. Unambiguous versions of three possible organizations of the subjective Necker cube: (A) the original ambiguous configuration; (B) cube in front; (C) cube in back; (D) a 'straddled' organization with the cube partly in front and partly in back

surface, however, its apparent brightness is considerably less. The same effect is present in comparing the brightness of the cubes in B and C of Figure 3, so it would seem that this difference in apparent brightness is not dependent on the presence of subjective as opposed to real contours.

Although the reader should be able to verify our observations simply by spending some time viewing Figure 1 (our experience indicates that these effects tend to emerge 'all or none'), it is, of course, necessary to demonstrate their existence in a naive sample.

METHOD

In Experiment I, 50 college students were individually shown a modified version of the subjective Necker cube in Figure 1.[2] In order to facilitate perception of the two main organizations, each observer was shown a 'preview' figure of an unambiguous cube-in-front or cube-in-back configuration. After being shown one of these preview figures for approximately 2 min, the observer viewed Figure 1 and was asked to describe what he saw. If the observer was able to see the cube in only one position, the other preview figure was used to prompt him as to where to look for the "other cube." Order of presentation of the preview figures was counterbalanced across observers. Finally, after the observer had achieved facility in perceiving both organizations of Figure 1, several forced-choice questions were asked, questions which required that the two cubes be compared to their backgrounds and to each other on various dimensions. The probing questions were always of the form "Which is brighter?" with the observer responding either "front" or "back" (for comparisons involving the cubes) or "background" or "cube" (for comparisons involving figure and ground).

In Experiment II, 35 college students viewed a slide of Figure 1. This experiment was conducted in a group situation.

RESULTS

Of the 50 observers tested in Experiment I, 1 was unable to perceive either of the possible arrangements of Figure 1, and 8 were unable to perceive the cube-in-front arrangement (each of these 8 had been shown the cube-in-back organization first). The remaining 41 observers were able to perceive the various aspects of the subjective Necker cube phenomenon: the cube-in-front organization with subjective contours extending between the discs, as well as the cube-in-back arrangement in which these same subjective contours are not present. Thus, 82% of the observers in the overall sample reported seeing the illusory bars of the cube in front [$p < .0001$, by the binomial test].[3] When asked to compare the relative brightness of the cube in front and the cube in back, 80% of the 41 observers perceiving both organizations reported the cube in

back as brighter [$p < .0001$]. Furthermore, for the cube-in-front organization the 'bars' of the subjective Necker cube were reported as brighter than the background for 68% of the 41 observers [$p < .027$]. It is important to note that this judgment was based on a comparison of the brightness of the bars *between* the discs (the subjectively seen portions of the cube) and the brightness of the background.

In all essential respects, the results of Experiment II replicated those of Experiment I. Again, the perception of subjective contours connecting the corners of the cube was present for the cube-in-front organization and absent for the cube-in-back organization for 91% of the observers [$p < .0001$]; the cube in back was perceived as brighter than the cube in front by 71% of the observers [$p < .012$]; and the cube in front was perceived as brighter than its background by 94% of the observers [$p < .0001$]. An additional finding was that 79% of the observers perceived the cube in back as being luminous [$p < .0004$], independently emitting its own light.

DISCUSSION

These data, though preliminary, do seem to substantiate the major perceptual effects of the subjective Necker cube described in the opening sections of this report. We would now like to briefly consider the implications of these findings for several current theories of subjective contour.[4]

First, since the location of the subjective contours seen in Figure 1 depends on the prevailing organization, it is unlikely (although not impossible) that peripheral physiological mechanisms such as lateral inhibition or simultaneous brightness contrast can account for the perception of subjective contour (Brigner and Gallagher, 1974; Frisby and Clatworthy, 1975). Recent data indicating that the apparent strength of subjective contour varies inversely (rather than directly, as predicted by contrast theory) with the illumination incident on the display is consistent with this conclusion (Dumais and Bradley, 1976).

Nor is it sufficient to say that subjective contour simply results from the presence of implicit depth cues which stratify one perceptual object in a plane over others — to argue, for example, that interposition cues in Figure 2 imply a triangular object partially occluding three black discs (Coren, 1972). Although depth information is certainly important in facilitating or inhibiting the perception of subjective contour (Lawson, Cowan, Gibbs, and Whitmore, 1974; Lawson and Gulick, 1967; Gregory and Harris, 1974; Harris and Gregory, 1973), this does not prove that depth cues per se are the causal factor. In cases where depth information is open to more than one interpretation, as in Figure 1 where the bars

of the cube can be seen as either overlying *or* underlying the white surface (B or C, Figure 3), the perceived location of the subjective contours will depend on how such ambiguous depth information is processed to organize objects stratified in depth (Bradley and Dumais, 1976).

Finally, recent attempts to account for subjective contour by assuming selective filtering of spatial frequencies (Ginsberg, 1975) would also seem to have difficulty explaining the alternative organizations of Figure 1, since a two-dimensional Fourier transform of the spatial information in that figure would presumably yield only one solution.

A theoretical perspective not yet considered by workers in the field involves subjective contour as the product of perceptual synthesis. In terms of a cognitive or information-processing theory of perception (Neisser, 1967; Hochberg, 1968, 1970), the visual system 'synthesizes' or 'constructs' a perceptual object under the guidance of both sensory and nonsensory information. Subjective contours represent an instance in which the visual system goes well beyond the information given in organizing sensory data, such that objects are created by interpolating edges and surfaces across objectively homogeneous regions of the visual field. Since the visual synthesis of a perceptual object is under both stimulus and central control, it may be influenced by all or some of the following: (a) directly given sensory information corresponding to the object, as in viewing a real cube or triangle; (b) contextual information conveyed by features in the stimulus array that indirectly imply the presence of a particular object, as in viewing Figure 2, where the sectors in the discs imply a masking triangle; (c) nonsensory factors like past experience, expectation, or perceptual set, as in being given an alternative construction of the cube in Figure 1. Where the available sensory information is particularly ambiguous, as in the subjective Necker cube, the relative importance of nonsensory factors in determining the visual synthesis is increased.

As is evident from Figure 3, there are at least three possible visual solutions to organizing the ambiguous array of Figure 1. Which particular solution is achieved would seem to depend primarily on central determinants rather than on stimulus factors, although peripheral factors such as eye movements during inspection would certainly be important. Given a particular visual solution, however, subjective contours are interpolated or 'filled in' throughout the visual field wherever necessary to demarcate the perceptual objects 'postulated' under the prevailing perceptual interpretation.[5] From this view, subjective contours are simply the phenomenal correlates of the edges or boundaries of visually constructed objects. Insofar as a given ambiguous array permits various possible constructions and

organizations of perceptual objects in depth, subjective contours will be seen in varying locations in the display (Bradley and Dumais, 1975).

Thus, when the cube-in-front inference is tentatively applied to the ambiguous array of Figure 1, the interpolation of contours over the intervening regions between discs occurs because *only this filling in will complete* and maintain the impression of an overlying cube. When the cube-in-back inference is adopted, however, the interpolation of contours over these same intervening regions would be incompatible with the requirement to maintain the impression of an occluding surface. Rather, the visual interpolation required in this instance gives rise to the illusory contours that complete the 'holes' of the occluding surface. Finally, the straddled-cube inference (D, Figure 3) requires a combination of the previous two forms of visual interpolation, with subjective contours extending between discs in some locations and completing holes in others.

In summary, the process of visual synthesis seems to produce that interpolation which would segregate perceptual objects into different planes and demarcate their phenomenal borders in such a way as to serve the prevailing perceptual interpretation. The pervasive effect of such organizational determinants of subjective contour argues against stimulus-bound or peripheralist explanations of the phenomenon.

Notes

The subjective Necker cube was discovered by the second author (while enrolled in a research seminar on subjective contour supervised by the first author), and he presented some of the data discussed in this report at the April, 1975, meeting of the Eastern Psychological Association. Support for this research was provided by a Faculty Research Grant (837-00-Bradley) awarded the first author by Bates College. Both authors thank Robert Moyer for a critical review of an earlier draft of this paper. Requests for offprints should be sent to Drake R. Bradley, Department of Psychology, Bates College, Lewiston, Maine 04240. Received for publication May 18, 1976; revision, June 10, 1976.

1. Although the connecting bars of the cube are not *visibly* present in the perception, they are phenomenally present as objects obscured by an occluding surface. This 'amodal' completion of the cube in back should be distinguished from the 'modal' completion of the cube in front, where the connecting bars are phenomenally *and* visibly present (this distinction comes from Kanizsa, 1974, 1976).

2. The display was chromatic with dark-blue inducing elements mounted on a bright-orange surface. It was viewed from a distance of about 12 ft so that the 'cube' subtended approximately 2.5 deg visual angle in lateral extent. Another experiment was conducted using an achromatic stimulus identical to Figure 1. The data indicated that the perceptual effects observed do not differ substantially as contingent on using chromatic or achromatic displays.

3. This test was based on $p = .50$ as chance expectation: that in response to the question "Do you see the bars of the cube extending between the discs?" half of the observers would respond 'yes' and half 'no,' even though illusory bars were never seen. This is certainly a conservative test, since the question asked the observer is highly unlikely to produce this strong a response bias (the appropriate level of expectation is $p = .00$ assuming no illusion and no effects due to response bias). Nevertheless, we must be alert to the possibility of response bias in these data, since some observers may just have reported seeing the illusory bars because of the expectation (perhaps created by the preview figure of the cube in front) that this is what they were 'supposed' to see. If so, no difference in the apparent brightness of the illusory bars and the background should be found, since that difference arises only with perceiving the subjective contour (i.e., the whiter-than-white subjective Necker cube). Since a significant proportion of the observers who reported seeing the illusory bars also reported seeing a difference in the apparent brightness of the illusory bars and the background, it would seem that for these individuals the illusory bars did have a phenomenal reality.

4. The theoretical significance of the difference in apparent brightness of the cube in front and the cube in back will not be discussed in this paper, as we wish to focus on the shifting subjective contours. However, we would note that the fact that the cube in back appeared brighter than the cube in front is consistent with Helmholtz's unconscious-inference theory of brightness constancy (Helmholtz, 1962). Since the cube in back appears to be situated in a dark enclosure of some sort, the amount of illumination incident on this cube would be 'registered' as less than that for the cube in front. However, since both cubes reflect the same absolute amount of light to the eye, the cube in back would therefore be judged as the intrinsically brighter object (see Coren and Komoda, 1973, and Hochberg and Beck, 1954, for related phenomena and theoretical interpretations). Further, the fact that the cube in back appeared luminous to some observers is also consistent with this theory. If the registered level of illumination in the cube-in-back enclosure is sufficiently low, then the actual amount of light reaching the eye from the cube may be greater than could be accounted for even by an object capable of reflecting 100% of the incident light. If so, the cube would then appear to be emitting its own light.

5. The view that perceptions represent 'hypotheses' selected by, but going beyond, sensory data has been emphasized by Gregory (1970, 1972). The role of nonsensory factors in determining the perceptual organization of ambiguous visual displays has been elaborated by Hebb (1949) in a neurophysiological theory capable of providing a physiological basis for the process of visual synthesis.

References

Bradley, D. R., and Dumais, S. T. 1975. Ambiguous cognitive contours. *Nature* 257:582–584.

Bradley, D. R., and Dumais, S. T. 1976. Simultaneous brightness contrast vs. organizational determinants of subjective contours. Submitted for publication.

Bradley, D. R., Dumais, S. T., and Petry, H. M. 1976. Reply to Cavonius. *Nature* 261:77–78.

Brigner, W. L., and Gallagher, M. B. 1974. Subjective contour: Apparent depth or simultaneous brightness contrast? *Perceptual and Motor Skills* 38:1047–1053.

Coren, S. C. 1972. Subjective contours and apparent depth. *Psychological Review* 79:359–367.

Coren, S. C., and Komoda, M. K. 1973. The effect of cues to illumination on apparent brightness. *American Journal of Psychology* 86:345–349.

Dumais, S. T., and Bradley, D. R. 1976. The effects of illumination level and retinal size on the apparent strength of subjective contours. *Perception and Psychophysics* 19:339–345.

Frisby, J. P., and Clatworthy, J. L. 1975. Illusory contours: Curious cases of simultaneous brightness contrast? *Perception* 4:349–357.

Ginsberg, A. P. 1975. Is the illusory triangle physical or imaginary? *Nature* 257:219–220.

Gregory, R. L. 1970. *The intelligent eye.* London: Weidenfeld and Nicolson.

Gregory, R. L. 1972. Cognitive contours. *Nature* 238:51–52.

Gregory, R. L., and Harris, J. P. 1974. Illusory contours and stereo depth. *Perception and Psychophysics* 15:411–416.

Harris, J. P., and Gregory, R. L. 1973. Fusion and rivalry of illusory contours. *Perception* 2:235–247.

Hebb, D. O. 1949. *The organization of behavior.* New York: Wiley.

Helmholtz, H. von. 1962. *Treatise on physiological optics,* vol. 3. New York: Dover.

Hochberg, J. 1968. In the mind's eye. In *Contemporary theory and research in visual perception,* ed. R. N. Haber. New York: Holt, Rinehart and Winston.

Hochberg, J. 1970. Attention, organization and consciousness. In *Attention: Contemporary theory and analysis,* ed. D. Mostofsky. New York: Appleton-Century-Crofts.

Hochberg, J., and Beck, J. 1954. Apparent spatial arrangement and perceived brightness. *Journal of Experimental Psychology* 47:263–266.

Kanizsa, G. 1955. Marzini quasi-percettivi in campi con stimolazione homogenea. *Rivista di Psicologia* 49:7–30.

Kanizsa, G. 1974. Contours without gradients or cognitive contours? *Italian Journal of Psychology* 1:93–113.

Kanizsa, G. 1976. Subjective contours. *Scientific American* 234(4):48–52.

Lawson, R. B., Cowan, E., Gibbs, T. D., and Whitmore, C. G. 1974. Stereoscopic enhancement and erasure of subjective contours. *Journal of Experimental Psychology* 103:1142–1146.

Lawson, R. B., and Gulick, W. L. 1967. Stereopsis and anomalous contour. *Vision Research* 7:271–297.

Neisser, U. 1967. *Cognitive psychology.* New York: Appleton-Century-Crofts.

Schumann, F. 1904. Einige Beobachtungen uber die Zusammenfassung von Gesichtseindrucken zu Einheiten. *Psychologische Studien* 1:1–32.

AMERICAN JOURNAL OF PSYCHOLOGY
June 1977, Vol. 90, No. 2, pp. 263–268

Instructions and grouped presentation in free recall

Pamela R. Sneed, David C. Brunts, and John H. Mueller
University of Missouri at Columbia

Subjects were presented either single items over trials for free recall, constant groupings of two words each over trials, increasingly larger groupings over trials, or progressively smaller groupings over trials. The nature of the grouped presentations had no effect on recall or organization; grouped presentation generally led to more organization than the single items, but the amount recalled was comparable in all cases. Instructions to use images led to more organized recall in some cases but increased recall only for the constant-size condition and hindered recognition somewhat in the changing-size conditions.

The present research was concerned with the effect of grouped presentation on free recall. Bower, Lesgold, and Tieman (1969, Experiment IV) used this method to test the hypothesis that recall of 'unrelated' words increases over trials partly because the subject's organizational units increase in size and integrity (e.g., Mandler, 1967; Tulving, 1968). On the assumption that simultaneous presentation of a subset of the list defines a useful organizational unit, Bower et al. structured over trials the sizes of the groupings of words presented for study. Subjects in an increasing-size condition were presented 3-word groupings on the first trial, 6-word groupings on the second trial, and 12-word groupings on the third. If subjects normally increase the size of their memory units over trials, this arrangement should facilitate recall. Subjects in a decreasing-size condition had the reverse sequence of groupings over trials. Bower et al. found greater recall and organization for the increasing-size condition, in apparent support of the 'chunking' interpretation of improved recall.

However, a recent report (Mueller and Overcast, 1975) questioned the use of the decreasing-size condition as a neutral baseline. The primary shortcoming is that if the increasing-size procedure *helps* subjects in their normal efforts to develop larger units, then the decreasing-size procedure would be expected to *hinder* those efforts. Thus, a more appropriate reference condition would be one that involves neither facilitation nor interference.

Mueller and Overcast (1975) used an arrangement with constant-size groupings over trials as the baseline condition in three experiments. Subjects in one constant-size condition were presented 3-word groupings each trial, and those in another had 12-word groupings each trial. In none of the three experiments did subjects in the increasing-size condition perform significantly better than those in the constant-size condition (in fact, in only one experiment did subjects in the increasing-size condition even perform better than those in the decreasing-size condition). Another reference condition had single-item presentation, to assess how much grouping per se contributed, as opposed to the *manner* of the grouping. The single-item condition showed less organization than the grouped conditions, but the amount recalled was comparable. The present study was a further investigation of the effects of grouped presentation, examining the effects of instructions and grouping size.

Bower et al. and Mueller and Overcast instructed all of their subjects to form images combining the words in a grouping presented for study. The intent of these 'imagery' instructions was to increase the likelihood that the groupings presented for study would serve as the functional memory units. However, it is possible that such instructions themselves induce sufficiently high levels of organization that effects due to changing the size of the groupings are masked. In other words, it seems possible that effects due to changes in the size of the study groupings might be more apparent if subjects were *not* told to form images during study. Therefore, half of the subjects in the present experiment were told to form an image uniting the words in a study grouping, while the other subjects received no such instructions.

Another factor examined in the present study was the absolute size of the study grouping. The smallest grouping used in the earlier studies was 3 words, with changes from 3 to 6 words and then 12 words over trials. Even if subjects normally incorporate more words into a chunk over trials, it is possible that such increments are tóo large to be useful. Therefore, the present study utilized groupings of 2, 4, and 8 words over trials, with the expectation that these smaller changes might be more manageable than was possibly the case in the earlier studies, thus maximizing any benefits of increasing the size of the groupings.

METHOD

Subjects

Eighty students from introductory psychology courses participated as subjects to fulfill course requirements. They were randomly assigned to the eight groups formed by the factorial combination of four *presentation conditions* (increasing-size, decreasing-size, constant-2, single-item study groupings) and

two types of *instructions* (general plus imagery instructions, general instructions only).

Materials

The word list was composed of 72 concrete nouns selected from the norms of Paivio, Yuille, and Madigan (1968). All had rated imagery greater than 5.10 and a Thorndike-Lorge frequency of at least 19 per million. They were randomly divided into 36 pairs for the 2-word study groupings. The pairs were then randomly combined to make the 4-word groupings, and these were then combined to make the 8-word groupings. The groupings were typed on 4-by-6-in. cards for presentation, with the original pairs maintaining their separate identity by appropriate spacing when larger groupings were formed.

The words within a pair also maintained the same spatial order on the card over trials, to minimize any confusion due to a change in their order. This was not done in Mueller and Overcast's study, and it is unclear whether it was done in the study by Bower et al., but it would seem to be the most favorable arrangement for demonstrating improvement due to increasing the size of the groupings.

Procedure

All subjects performed for three free-recall trials. Those in the *increasing-size* condition received study groupings of 2 words each on the first trial, 4 words each on the second trial, and 8 words each on the third. Those in the *decreasing-size* condition had the reverse sequence of sizes over trials. Those in the constant-size condition (*constant-2*) received the same groupings of 2 words each on every trial, with only the order of presentation of groupings varying. Subjects in all of these groups were told that the size of the study grouping might change or stay the same over trials. In each of these conditions, the multiple-word groupings were shown for a period of time that allowed 4 sec per word (8 sec, 16 sec, or 32 sec per card). Subjects in the *single-item* condition had the list presented one word at a time by a slide projector, at a 4-sec rate, with the order of the words changing over trials. After each complete presentation of the list, subjects had a 3-min written test of recall.

In addition to *general* instructions that emphasized the option to recall in any order, half of the subjects in each of the four presentation conditions received additional *imagery* instructions suggesting that they form mental images combining the words on a card as interacting in some way, with an illustrative example.

The third test of recall was followed by an unpaced test of recognition. Subjects were given a list of 144 words, the 72 old words plus 72 new words, and required to identify the words they had seen before in the experiment.

RESULTS

Recall

The average recall per trial is shown in Table 1 for each group. The only significant effect was the main effect of trials [$F(2, 144) = 507.75$, $MS_e = 19.52$], as learning occurred in all groups.[1] No other differences were significant [$Fs < 1.92$].

Table 1. Average recall per trial in each condition

	Presentation condition			
	Increasing	Decreasing	Constant-2	Single-item
Imagery instructions	30.43	30.13	36 87	35.03
No imagery instructions	33.57	34.33	33.57	35.37

Organization

Recall was scored for intertrial *repetitions* of 2 words, 4 words, and 8 words, these being the grouping sizes shown to subjects. An unordered criterion for repetitions was used, and Pellegrino's (1972) adjusted ratio of clustering (*ARC*) score was computed in each case. The average scores are shown in Table 2, averaged over trials.

Imagery instructions led to more organization in general [$F(1, 72) = 3.49, p < .07, MS_e = .02$], and all grouped presentation conditions showed more organization than the single-item condition [$F(3, 72) = 6.29, MS_e = .02$]. The interaction of instructions and presentation condition was marginally significant [$F(3, 72) = 2.54, p < .07, MS_e = .02$], as imagery instructions improved organization in the constant-2 and decreasing-size conditions but not in the increasing-size and single-item conditions. These effects were most apparent for intertrial repetitions of 2 words (see Table 2), as indicated by significant interactions of instructions by repetitions [$F(2, 144) = 5.54$], presentation condition by repetitions [$F(6, 144) = 7.01$], and instructions by presentation condition by repetitions [$F(6, 144) = 2.39$; all $MS_e = .01$].

Organization increased over trials [$F(1, 72) = 7.49, MS_e = .01$], but

Table 2. Average adjusted ratio of clustering score

Presentation condition	Intertrial repetitions		
	2 words	4 words	8 words
Imagery instructions			
Increasing	.39	.06	.00
Decreasing	.48	.08	.00
Constant-2	.50	.10	.00
Single-item	.25	.01	.00
No imagery instructions			
Increasing	.42	.05	.00
Decreasing	.31	.06	.01
Constant-2	.37	.04	.00
Single-item	.24	.07	.00

neither instructions nor presentation condition interacted with trials [$Fs < 1.08$]. There was a marginally significant triple interaction of instructions by presentation condition by trials [$F(3, 72) = 2.62, p < .06, MS_e = .01$]. This interaction indicated that the increase in organization over trials occurred in all subgroups, except for the decreasing-size condition with imagery instructions and the constant-2 condition without imagery instructions (both of which showed a decline over trials) and the single-item condition with imagery instructions (which remained virtually the same over trials). These effects were also greatest for intertrial repetitions of 2 words. The interaction of repetition size by trials was not significant [$F(2, 144) = 1.95$], as the number of repetitions of all sizes increased over trials. The interaction of presentation condition by repetitions by trials was not significant [$F(6, 144) = 1.97$], as would have been the case if subjects under one presentation condition were utilizing larger chunks over trials.

Recognition

Table 3 presents the average number of correct recognitions of old items for each subgroup. There were no main effects of instructions or presentation condition [$Fs < 1.12$]. The interaction of instructions by presentation condition was marginally significant [$F(3, 72) = 2.25, p < .09, MS_e = 19.44$], as imagery instructions increased recognition in the constant-2 and single-item conditions but reduced it in the increasing-size and decreasing-size conditions.

DISCUSSION

These results fail to support the hypothesis that larger study groupings over trials will facilitate performance. Recall was not improved by the increasing-size arrangement relative to either decreasing-size or constant-2 conditions. If anything, both increasing-size and decreasing-size arrange-

Table 3. Average correct recognitions in each condition

	Presentation condition			
	Increasing	Decreasing	Constant-2	Single-item
Imagery instructions	67.70	67.40	69.70	70 50
No imagery instructions	69.00	69.80	65.40	69.60

Note: Maximum score, 72.

ments led to somewhat worse performance than the constant-2 condition, as if the subjects under both former conditions experienced some interference due to the changing composition of study groupings over trials. This was the case for both types of instructions, and the pattern of results was much the same for recall and subjective organization. The grouped presentation conditions did show more organization than the single-item condition, but this advantage did not carry over to recall.

Thus, the previous failure to replicate the difference between increasing-size and decreasing-size conditions does not seem attributable to a change in the spatial order of the words in groupings over trials. Nor does the difference seem related in any simple manner to either imagery instructions or the absolute number of words in the study unit. Variations in the manner of sequencing subsets of different sizes might conceivably have some effect with nominal intralist relationships (e.g., instances of taxonomic categories) but would seem to be of little consequence with 'unrelated' words, since subjects seem to be able to compensate for most such variations.

Notes

Requests for offprints should be sent to John Mueller, Psychology Department, 210 McAlester, University of Missouri, Columbia, Missouri 65201. The authors acknowledge the comments by Donald Kausler on an earlier draft of this paper. Received for publication June 17, 1976.

1. Effects described as significant involve $p < .05$ or better, unless noted otherwise.

References

Bower, G. H., Lesgold, A. M., and Tieman, D. 1969. Grouping operations in free recall. *Journal of Verbal Learning and Verbal Behavior* 8:481–493.

Mandler, G. 1967. Organization and memory. In *Psychology of learning and memory,* vol. 1, ed. K. W. Spence and J. T. Spence. New York: Academic Press.

Mueller, J. H., and Overcast, T. O. 1975. Grouped presentation in free recall. *American Journal of Psychology* 88:649–660.

Paivio, A., Yuille, J. C., and Madigan, S. A. 1968. Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monograph* 76 (1, part 2).

Pellegrino, J. W. 1972. A FORTRAN IV program for analyzing higher-order subjective organization units in free recall learning. *Behavior Research Methods and Instrumentation* 4:215–217.

Tulving, E. 1968. Theoretical issues in free recall. In *Verbal behavior and general behavior theory,* ed. T. R. Dixon and D. L. Horton. Englewood Cliffs, N.J.: Prentice-Hall.

AMERICAN JOURNAL OF PSYCHOLOGY
June 1977, Vol. 90, No. 2, pp. 269-280

Frequency and meaningfulness in tachistoscopic word perception

Leon Manelis
Illinois State University

This study investigated the superior perceptibility of words over regular, pronounceable nonwords in tachistoscopic displays. Paradoxically, this effect was demonstrated in the absence of a word-frequency effect. The results suggest that the superior perceptibility of words in tachistoscopic displays is due to highly specific characteristics of letters as they occur in words and that lexical retrieval is not involved.

In studies on the tachistoscopic perception of letters and strings of letters, a major concern is the role of linguistic variables. There are numerous demonstrations that words and other strings that are regular in spelling and pronunciation are perceived more accurately than strings of letters that are not regular (see reviews by Krueger, 1975; Massaro, 1975; Smith and Spoehr, 1974). These linguistic effects are eliminated under some testing conditions, especially when subjects are given a very small set of target letters and are required to detect one of the targets in a string (Estes, 1975a, 1975b). However, under a broad range of other testing conditions, the linguistic effects are reliable.

The purpose of this paper is to consider one of the linguistic effects, one elsewhere termed the *meaningfulness* effect (Manelis, 1974). This refers to the superior perceptibility of tachistoscopically presented words over regularly spelled, pronounceable nonwords. The earlier paper was in response to an initial failure to find a meaningfulness effect (Baron and Thurston, 1973), but the paper reported experiments that did demonstrate it. The effect has also been obtained by Juola, Leavitt, and Choe (1974) and by McClelland (1976, Experiment I). Spoehr and Smith (1975) replicated the effect in one experiment, though not in another. The present paper reports three more experiments that demonstrate the effect, thus providing further support for its reliability. The main concerns, though, are with the theoretical implications of the new evidence.

There have been two distinguishable theoretical approaches to the problem of word perception. One approach is to investigate the structural regularities of spelling and pronunciation that characterize words, with no special interest in the process of lexical retrieval (e.g., Gibson, 1965; Spoehr and Smith, 1975). The other approach centers on lexical retrieval (e.g., Broadbent and Broadbent, 1976; Morton, 1969). There are also theoretical discussions that combine lexical and structural features (e.g., Rubenstein, Lewis, and Rubenstein, 1971b; Rumelhart and Siple, 1974). The meaningfulness effect seems to imply the occurrence of lexical retrieval in tachistoscopic perception — to suggest that words, but not even the most regularly structured nonwords, are encoded as familiar categories. If lexical retrieval is indeed the explanation of the meaningfulness effect, one would expect word *frequency* to have an effect as well. Experiments I and II tested this prediction.

EXPERIMENT I

It has been shown that word frequency is an effective variable in lexical retrieval. In the lexical-decision task, which requires subjects to decide whether verbal items are words or nonwords, it has been found that common words are responded to faster than uncommon words (Rubenstein, Garfield, and Milliken, 1970; Rubenstein, Lewis, and Rubenstein, 1971a, 1971b). A similar result has been found in a naming task (Forster and Chambers, 1973). The latency for pronouncing common words was less than that for uncommon words; pronounceable nonwords had still longer latencies than the words. These findings suggest that a common word can be accessed in lexical memory faster than an uncommon word.

The results from the lexical-decision and naming tasks have relevance for the tachistoscopic task. Because common words can be accessed faster than uncommon words, the brief time that a tachistoscopic display is available should more often permit successful lexical retrieval of common words than of uncommon words. Thus, if lexical retrieval is involved in tachistoscopic perception, common words should be perceived more accurately than uncommon words.

There has been previous research demonstrating frequency effects in tachistoscopic recognition (e.g., Newbigging, 1961). However, a problem with the previous research is the use of free report: the subjects simply reported the word they thought was presented on each trial. It is known that when subjects are required to guess a word with only one or two letters as a cue, they produce more common words than uncommon words (Duncan, 1966). Similarly, it may be that in tachistoscopic recognition,

there is a bias for reporting common words in response to the incomplete information from the brief display. This response bias might produce an apparent frequency effect.

To eliminate the requirement that the subjects produce a word as a response, a forced-choice test may be used. In the present experiment, words and pronounceable nonwords were presented tachistoscopically, and a forced-choice test with two alternatives followed each presentation (Reicher, 1969). The alternatives precluded the influence of biases in subjects' choices, because either alternative (the correct choice or the distractor) formed an item of the type presented tachistoscopically: a pronounceable nonword or a word at a particular level of frequency.

METHOD

Stimuli

The stimuli tested were 96 words and 96 pronounceable nonwords. All had four letters and were of one syllable. The *words* were selected to form pairs in which both members were at the same level of frequency and each member differed from the other by only a single letter (e.g., COLD, HOLD). The two letters by which the members of each pair differed were the alternatives that were presented for test after the tachistoscopic display of either member; thus, the test was a probe of a single letter position. Corresponding to each pair of words was a pair of pronounceable *nonwords* that were formed from the words by changing one letter (e.g., CULD, HULD). The changed letter was always at a position other than the probe. The contrasting letters for the nonword pair were thus the same as for the matched word pair, and they served as the test alternatives for the nonwords as well as for the words. There were 16 stimulus quartets (pairs of words and matched pairs of nonwords) at each of three levels of frequency (Kučera and Francis, 1967). The low-frequency words ranged from 2 to 10 occurrences per million, with a median of 3. The medium-frequency words ranged from 23 to 74 occurrences per million, with a median of 37. The high-frequency words ranged from 94 to 895 occurrences per million, with a median of 211. Within each level of frequency, one-fourth of the stimuli were probed at each letter position.

Materials

The stimuli were typed on white cards with an IBM Executive Registry typewriter; they subtended a visual angle of .2 deg vertically and .9 deg horizontally. The test cards each contained four dashes in a row, corresponding to the four letters of the stimuli. Above and below the dash representing the probe position were the two alternative letters. This array subtended a visual angle of .6 deg vertically and 1.4 deg horizontally.

Design and procedure

Words and nonwords were randomly intermixed. One word and one nonword from each stimulus quartet were presented at each of two testing sessions. At each

session, 48 words and nonwords different from those used in the main experiment were presented as warm-up trials, during which the exposure duration was adjusted with the aim of achieving 75% accuracy for each subject.

The materials were presented in a three-field Iconix tachistoscope. At the outset of each trial, the subject viewed a fixation field, consisting of two horizontal gray lines bounding the area where the stimuli would appear. After the experimenter's 'ready' signal, the subject pressed a button to initiate the trial. Half a second after the button was pressed, the fixation field went off and the stimulus appeared for the adjusted exposure duration. The stimulus was immediately followed by a mask (a crosshatched pattern) in the same area, and the two letters that were the alternatives above it. The subject chose between the alternatives by pressing one of two buttons. A response was required on all trials, but there was no time constraint. When either response button was pressed, the mask and alternatives went off and the fixation field reappeared.

Subjects

Twenty psychology students served as subjects. All were native speakers of English and served for two sessions of about 45 min each.

RESULTS AND DISCUSSION

The proportions correct are shown in Table 1, broken down by level of frequency for the words and their matched nonwords. Statistical analysis was performed by treating both subjects and items as random effects (Clark, 1973; Winer, 1971, p. 375). This technique allows generalization of the results to both a population of subjects and a population of items within the constraints of selection. There was a significant effect of meaningfulness [$min\ F'(1, 52) = 8.71, p < .005$]. But neither the effect of frequency nor the interaction of frequency by meaningfulness was significant

Table 1. Results of Experiments I and II: Word frequency

Experiment I

	Frequency		
	Low	Medium	High
Words	.775	.794	.800
Nonwords matched to words	.736	.744	.747

Experiment II

		Matched words	
	Nonwords	Common	Uncommon
	.706	.762	.757

[*min F*'s < 1]. A separate analysis was performed on the words alone, and again the effect of frequency was not significant [*min F*' < 1]. Nor was the effect of frequency significant in separate analyses of variance that treated subjects or items, but not both, as random effects [*F*s < 1]. Thus, although the meaningfulness effect was replicated, there was no evidence of a frequency effect.

Accepting a null hypothesis always involves some risk. In the present data, moreover, there was for the words a slight trend toward increasing accuracy with increasing frequency. However, there was a similar trend for the matched *nonwords*. This is not surprising, because the test letters and most of the remaining letters in the stimuli were matched only between words and nonwords *within* each level of frequency, not across levels of frequency. The trend for nonwords makes it difficult to accept the trend for words as an effect of lexical retrieval.

In the naming task discussed above (Forster and Chambers, 1973), the difference in reaction times between common and uncommon words was much greater than that between uncommon words and nonwords. According to a lexical-retrieval hypothesis, then, one might expect the difference across levels of frequency in the present experiment to be greater than the difference between words and nonwords, but this was not so. (The range of frequency in the two experiments is nearly identical. A similar comparison of effects from the lexical-decision task is not possible because the words produce positive responses, whereas the nonwords produce negative responses.)

A further way to assess the conclusion that there was no frequency effect is to do another experiment and test for the effect again. This was the purpose of Experiment II.

EXPERIMENT II

Two major changes were made from Experiment I in order to provide a more powerful test for a frequency effect. First, words and nonwords were separately blocked rather than randomly intermixed. It may be that lexical access is more likely to occur when words and nonwords can be expected consistently from trial to trial. If so, then a frequency effect should also be more likely when words are blocked together. The second change from Experiment I was that high- and low-frequency words were closely matched on an individual basis. Sets of high- and low-frequency words were selected with the same letters at the probe positions and with as many of the same letters as possible at the remaining positions (e.g., high-frequency words, PART, PARK; low-frequency words, PERT, PERK).

METHOD

Stimuli

The stimuli were selected to form sets of six items: a pair of high-frequency words, a pair of low-frequency words, and a pair of nonwords. Within each pair, the two members differed from each other by one letter. The contrasting letters served as the test alternatives, and they were the same (also at the same position) for the three pairs of each set. After the stimulus sets had been selected, the question arose as to whether some of the low-frequency words would be known as words by the subjects (e.g., PERT). Accordingly, ratings were obtained from 20 introductory psychology students for the words and nonwords that had been selected, as well as some additional items to be used as possible replacements. The words and nonwords were randomly intermixed and listed on two sheets of paper. Subjects were instructed to decide whether each item was a word or not and to rate their confidence in their decisions on a scale from 1 to 4. After inspection of these data, two words and one nonword were replaced. In the final stimulus set, there were 16 pairs of common or high-frequency words (frequency at least 20 per million; median, 119), 16 pairs of uncommon or low-frequency words (less than 20 per million; median, 5), and 16 pairs of nonwords. In the final set of items, there were three failures to recognize the high-frequency words as words and three failures to recognize the low-frequency words: the mean confidence ratings were 4.00 and 3.96 respectively for correct recognitions. The mean number of false recognitions for the nonwords was 2.57 (out of 32 nonwords), and on the scale from 1 to 4, the mean confidence ratings were 1.76 for false recognitions and 3.06 for correct rejections of the nonwords.

Design and procedure

The words were divided into two blocks of 32 trials, each containing high- and low-frequency words. The nonwords were grouped into a third block. The sequence of the three blocks was counterbalanced across subjects. For half of the nonword pairs, a filler item was added to the nonword block in order to discourage the possibility of a certain strategy. Because both members of each pair of nonwords were presented in the same block (e.g., RABE, LABE), the same test alternatives (R, L) were presented twice. To discourage subjects from responding to the second occurrence simply by choosing the letter that was wrong on the first presentation, the filler item was included with the same alternatives presented again. The equipment, materials, and procedure were the same as in Experiment I, except that reaction time was also recorded.

Subjects

Twenty-four psychology students served as subjects; all were native speakers of English. Each subject served for one session of about 50 min.

RESULTS AND DISCUSSION

The proportions correct for nonwords, uncommon words, and common words are shown in Table 1. Planned orthogonal comparisons among the

three means indicated a significant difference between words and nonwords but no difference between common and uncommon words [respectively, *min* $F'(1, 65) = 4.65$, $p < .05$, and *min* $F' < 1$]. Nor was the effect of frequency significant in separate analyses that treated subjects or items, but not both, as random effects [$Fs < 1$]. Mean reaction times to correct responses were 1,339, 1,328, and 1,317 msec for nonwords, uncommon words, and common words respectively. Statistical analysis indicated no significant differences [*min* F's < 1].

Thus, both Experiments I and II show that a meaningfulness effect was obtained without a frequency effect. This finding is somewhat paradoxical: familiarity with the stimuli that were words enabled them to be perceived more accurately than the nonwords, but the different degrees of familiarity for common and uncommon words had no effect. This conclusion is relevant to the earlier discussion of lexical access. On the basis of the findings from the lexical-decision and naming tasks, it was argued that a frequency effect should be obtained in the tachistoscopic task. The assumption in predicting the frequency effect was that the limited stimulus exposure restricts the probability of lexical retrieval. (Smith and Spoehr, 1974, make the similar assumption that exposure duration limits the time available for completing stages of processing.) However, one might object that the stimulus is not necessarily encoded as a word while physically present; instead, perhaps the subject derives a certain amount of information from the stimulus during its display and lexically encodes that information after the display has ended. There is no time limit on processing after the display, and the extra time needed to access an uncommon word as a code for the stimulus information would not affect accuracy. For nonwords, of course, an appropriate lexical code could not be retrieved at all, and perhaps this failure would explain the general meaningfulness effect.

On closer examination, however, this failure in itself would not be an adequate explanation. Instead, it is necessary to consider a decay process for the stimulus information. If the unencoded stimulus information extracted from the display were fully retained when the subject decided between the two alternative letters, no meaningfulness effect would be obtained, because words and pronounceable nonwords would not differ on the basis of that unencoded information. Lexical encoding would have an effect only if it preserved the otherwise decaying stimulus information. But because the decaying stimulus information must be available in order for encoding to occur, relatively slow lexical retrieval (for uncommon words as opposed to common words) would be less likely to succeed. This line of reasoning leads again to the conclusion that a frequency effect

should be obtained. Thus, the argument that lexical retrieval does occur, but only after the stimulus display, is not correct. A further disconfirmation is based on the reaction times measured in Experiment II. If lexical retrieval were occurring, one would expect responses to uncommon words to be slower than responses to common words, but there was no difference.

If lexical retrieval does not then determine accuracy in tachistoscopic perception, what explains the meaningfulness effect? The alternative to lexical retrieval is that perception is tuned to the specific structure of words — to some characteristic of the letters in words that allows them to be perceived more accurately than nonwords. A convincing demonstration that this characteristic had been identified would be, first, to select words and nonwords so that the nonwords satisfy the characteristic more than the words and, then, to find that processing of the nonwords is superior to processing of the words. Mason's study (1975) is the only one that has provided such a demonstration. The characteristic of letters that Mason investigated is their positional frequency, the frequency with which letters occur in words at particular positions (Mayzner and Tresselt, 1965). The task in her study was searching for a target letter in a visual display. In some cases, good readers were faster at this task when the display consisted of nonwords of high positional frequency than when it consisted of words of lower positional frequency. Although Mason's task was not tachistoscopic perception, the unique reversal of the usually superior performance on words recommends positional frequency as the most likely variable to consider in an attempt to explain the meaningfulness effect. The purpose of Experiment III was to provide the necessary test of this variable.

EXPERIMENT III

Because positional-frequency counts are based on real words, one can generally expect nonwords to have lower positional frequency. However, it is possible to select nonwords with higher positional frequency. If positional frequency is a characteristic of strings of letters that accounts for the meaningfulness effect, then the effect should be reversed with these specially selected items. The present experiment tested this prediction.

METHOD

Stimuli and design

The stimuli were 48 words and 48 nonwords, in sets of four. All stimuli were four letters long, were pronounceable, and were presented twice to each sub-

ject. Within each set of four stimuli were two words and two nonwords, all of which differed from each other at only one letter position. At this position, the letters that formed words had relatively low positional frequency, with a median of 72 out of 20,000 words, and the letters that formed nonwords had high positional frequency, with a median of 389 (Mayzner and Tresselt, 1965). For each set of items, pairs of the contrasting letters served as test alternatives. In condition W(W), after a word was presented tachistoscopically, the test alternatives were one of the original letters (the correct alternative) and the letter that formed the other word in the set (incorrect alternative). In condition W(N), the tachistoscopic display was a word, and the incorrect test alternative was a letter that formed a nonword in the set. In condition N(N), the tachistoscopic display was a nonword, and the incorrect alternative was the letter that formed the other nonword in the set. In condition N(W), the tachistoscopic display was a nonword, and the incorrect alternative was a letter that formed a word in the set. Across the four conditions, the confusability of the correct and incorrect alternatives was matched using Townsend's (1971) data, which are based on a type font very similar to the letters used in the present experiment. All four conditions were intermixed within the sequence of trials.

Materials and procedure

The materials and procedure were the same as in Experiment I, with one exception. At the outset of each trial, subjects were told to fixate a marked position corresponding to the letter position that would be tested. This procedure was based on the assumption that positional frequency would most likely be effective if subjects could focus on the letters whose frequency at that position was being manipulated.

Subjects

Twenty psychology students served as subjects. Each participated for two sessions of about 50 min.

RESULTS AND DISCUSSION

The proportions correct are shown in Table 2. The statistically significant finding was that words were perceived more accurately than nonwords [$min\ F'(1, 34) = 6.04$, $p < .025$]. Clearly, the meaningfulness effect was not reversed, demonstrating that despite the suggestion of Mason's (1975) data, positional frequency is not responsible for this effect.

Again, Experiments I and II produced the seemingly paradoxical result

Table 2. Results of Experiment III: Positional frequency

Words		Nonwords	
W(W)	W(N)	N(N)	N(W)
.714	.722	.650	.651

of a meaningfulness effect in the absence of a frequency effect. As discussed above, if lexical retrieval were to account for the meaningfulness effect, a frequency effect should also occur. Thus, the meaningfulness effect cannot be explained by lexical retrieval. The alternative explanation (and now the favored one) is that the processes for perceiving strings of letters operate in a way that is sensitive to the unique characteristics of the letters in words. In order to account for the meaningfulness effect, these characteristics must differ for words on the one hand and apparently regular, pronounceable nonwords on the other hand. The perceptual processes involved must be finely tuned and sensitive to some subtle differences. Mason's (1975) findings had suggested that positional frequency is a characteristic likely to distinguish words and regular nonwords, but Experiment III showed that it does not account for the meaningfulness effect. Nor can translation models account for the effect, because they explain the perception of letters in terms of regular, phonological rules that are not specific to words but apply equally well to regular nonwords. Thus, a shortcoming of Spoehr and Smith's otherwise highly successful translation model (1973, 1975) is its failure to account for the meaningfulness effect: words and regular nonwords do not differ in terms of the relevant variables that Spoehr and Smith have identified, variables such as syllables or patterns of consonants and vowels.

There have been other studies involving brief visual presentation or noisy auditory presentation of words, and many of them have demonstrated frequency effects (e.g., Broadbent and Broadbent, 1975; Rumelhart and Siple, 1974). An important difference between those studies and the present one is that they required subjects to identify whole words, whereas here subjects were tested only on single letters. Lexical retrieval may be especially likely when subjects must identify complete items; in such a situation, a frequency effect might well be found. Clearly, lexical retrieval must occur in some situations, as in normal reading. But the present results imply that because of structural characteristics that are specific to words, they are perceived more accurately than nonwords, even in the *absence* of lexical retrieval.

Notes

The author thanks David Tharp for assistance in data collection and analysis. This research was supported in part by Grant 1 R03 MH28068-01 from the National Institute of Mental Health. Requests for offprints should be sent to Leon Manelis, Department of Psychology, Illinois State University, Normal, Illinois 61761. Received for publication August 10, 1976.

References

Baron, J., and Thurston, I. 1973. An analysis of the word-superiority effect. *Cognitive Psychology* 4:207–228.

Broadbent, D. E., and Broadbent, M. H. P. 1975. Some further data concerning the word frequency effect. *Journal of Experimental Psychology: General* 104:297–308.

Clark, H. H. 1973. The language-as-fixed effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior* 12:335–359.

Duncan, C. P. 1966. Effect of word frequency on thinking of a word. *Journal of Verbal Learning and Verbal Behavior* 5:434–440.

Estes, W. K. 1975a. The locus of inferential and perceptual processes in letter identification. *Journal of Experimental Psychology: General* 104:122–145.

Estes, W. K. 1975b. Memory, perception, and decision in letter identification. In *Information processing and cognition: The Loyola Symposium*, ed. R. L. Solso. Potomac, Md.: Lawrence Erlbaum Associates.

Forster, K. I., and Chambers, S. M. 1973. Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior* 12:627–635.

Gibson, E. J. 1965. Learning to read. *Science* 148:1066–1072.

Juola, J. F., Leavitt, D. D., and Choe, C. S. 1974. Letter identification in word, nonword, and single letter displays. *Bulletin of the Psychonomic Society* 4:278–280.

Krueger, L. E. 1975. Familiarity effects in visual information processing. *Psychological Bulletin* 6:949–974.

Kučera, H., and Francis, W. N. 1967. *Computational analysis of present-day American English*. Providence, R.I.: Brown University Press.

Manelis, L. 1974. The effect of meaningfulness in tachistoscopic word perception. *Perception and Psychophysics* 16:182–192.

Mason, M. 1975. Reading ability and letter search time. Effects of orthographic structure defined by single-letter positional frequency. *Journal of Experimental Psychology: General* 104:146–166.

Massaro, D. W. 1975. Primary and secondary recognition in reading. In *Understanding language*, ed. D. W. Massaro. New York: Academic Press.

Mayzner, M. S., and Tresselt, M. E. 1965. Tables of single-letter and diagram frequency counts for various word-length and letter-position combinations. *Psychonomic Monograph Supplements 1* (2).

McClelland, J. L. 1976. Preliminary letter identification in the perception of words and nonwords. *Journal of Experimental Psychology: Human Perception and Performance* 2:80–91.

Morton, J. 1969. Interaction of information in word recognition. *Psychological Review* 76:165–178.

Newbigging, P. L. 1961. The perceptual redintegration of frequent and infrequent words. *Canadian Journal of Psychology* 15:123–132.

Reicher, G. M. 1969. Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology* 81:274–280.

Rubenstein, H., Garfield, L., and Milliken, J. A. 1970. Homographic entries in the internal lexicon. *Journal of Verbal Learning and Verbal Behavior* 9:487–494.

Rubenstein, H., Lewis, S. S., and Rubenstein, M. 1971a. Homographic entries in the internal lexicon: Effects of systematicity and relative frequency of meanings. *Journal of Verbal Learning and Verbal Behavior* 10:57–62.

Rubenstein, H., Lewis, S. S., and Rubenstein, M. 1971b. Evidence for phonemic recoding in visual word recognition. *Journal of Verbal Learning and Verbal Behavior* 10:645–657.

Rumelhart, D. E., and Siple, P. 1974. The process of recognizing tachistoscopically presented words. *Psychological Review* 81:99–118.

Smith, E. E., and Spoehr, K. T. 1974. The perception of printed English: A theoretical perspective. In *Human information processing: Tutorials in performance and cognition*, ed. B. H. Kantowitz. Potomac, Md.: Lawrence Erlbaum Associates.

Spoehr, K. T., and Smith, E. E. 1973. The role of syllables in perceptual processing. *Cognitive Psychology* 5:71–89.

Spoehr, K. T., and Smith, E. E. 1975. The role of orthographic and phonotactic rules in perceiving letter patterns. *Journal of Experimental Psychology: Human Perception and Performance* 1:21–34.

Townsend, J. T. 1971. Theoretical analysis of an alphabetic confusion matrix. *Perception and Psychophysics* 9:40–50.

Winer, B. J. 1971. *Statistical principles in experimental design*. New York: McGraw-Hill.

AMERICAN JOURNAL OF PSYCHOLOGY
June 1977, Vol. 90, No. 2, pp. 281-289

Recall as a function of the relationship between the cues and the words to be remembered

Ruth H. Maki
North Dakota State University

Superordinate cues (e.g., 'animal' for 'dog') and coordinate cues (e.g., 'cat' for 'dog') were compared in two experiments. In Experiment I, superordinate cues produced both higher recall and better guessing of the words to be recalled than did coordinate cues. In Experiment II, coordinate and superordinate cues were selected so that they were equally likely to elicit the words to be recalled as free associates; here the two types of cues were equally effective in recall and guessing. Associability and not the superordinate or coordinate relationship seems to be important in determining the effectiveness of cues.

Tulving (1972) has distinguished between two types of memory. Episodic memory is studied in traditional list-learning experiments in which a word is marked as having been presented in a given context. Semantic memory consists of an individual's body of knowledge, the organization of which has usually been studied with reaction time as the dependent variable in sentence-verification tasks. Very little research has been addressed to the role in episodic memory of variables important in semantic memory. The experiments presented in this paper investigated the role of links between superordinates and instances of them, an important relationship in semantic memory, in episodic memory. The predictions made in these experiments were based largely on Quillian's (1966) model of semantic memory, a model recently expanded by Collins and Loftus (1975). This model stresses the importance of links between superordinates and instances of them in a semantic network. When a subject is making a decision about whether 'an X is a Y,' such links have strong positive effects, although this is not the only type of evidence involved in the decision. If these links are important in such semantic-memory tasks, such relationships might also be important in an episodic-memory task, such as free recall of word lists.

Anderson (1972) compared superordinates as cues for words (e.g., 'animal' for 'dog') with cues assumed to be at the same level in a hierarchy as the words to be remembered (e.g., 'cat' for 'dog'), a relationship

which will be called coordinate in this paper. He presented sentences (e.g., 'The bungalow stood near the river') and asked for recall of the last word in the sentence, which in this example would be 'river.' Either the subject of the sentence was itself presented as a cue or a superordinate close to that subject in the assumed hierarchy (e.g., 'dwelling' for 'bungalow') or more remote in the hierarchy (e.g., 'building'). Likewise, coordinate cues were either close in the hierarchy (e.g., 'cabin') or remote in the hierarchy (e.g., 'cathedral').

Anderson found that the subjects of the sentences were the best cues for recall of the last word in the sentence and that superordinate cues produced better recall than did coordinate cues under both conditions, remote and close. He interpreted his results within the framework of Quillian's hierarchical model of semantic memory. A cue was assumed to be the starting point of a search through semantic memory, a search which begins in nodes below the cue. When the cue is a superordinate, the subject of the sentence is located directly beneath it for a close superordinate and farther beneath it for a remote superordinate. However, in both cases search below the cue will contact the subject of the sentence. When a cue is a coordinate, a match will not be found in nodes below the cue. One must move up a node and then search below that higher node. A match would then be found for close coordinates, but one must go up still another node in order to find a match for a remote coordinate. Since, on the average, more nodes must be searched with coordinate cues, the probability of locating the subject of the sentence is lower than with superordinate cues. Anderson ruled out the frequency of the words' usage in the language and their concreteness as factors in the superiority of superordinate cues and concluded that the superordinate relationship of the cue to the subject of the sentence was critical in recall.

A somewhat different interpretation of Anderson's results involves Tulving's (1972) distinction between episodic and semantic memory. Tulving has hypothesized that words presented for recall are placed in episodic memory. Cues presented during learning or related words from semantic memory may be part of the encoding in episodic memory. Since the cues were not presented during learning in Anderson's study, the probability of a cue being stored with a word in episodic memory would be a function of the relationship in semantic memory. Superordinates might be more likely to be part of a word's encoding in episodic memory than coordinates, since the category membership of a word is an important part of its meaning. In his encoding-specificity hypothesis, Tulving (Thomson and Tulving, 1970; Tulving and Osler, 1968) has argued that only cues in episodic memory can aid retrieval of an item during list recall. Therefore, super-

ordinates, which are likely to be entered in episodic memory, would be better cues than coordinates, which are less likely to be in episodic memory.

The encoding-specificity hypothesis also makes predictions about cued recall when *input* is cued. Although Tulving has tested the encoding-specificity hypothesis by comparing various types of cued recall under constant conditions of input, predictions about cued recall with the two types of cues present or absent at input follow from his views about episodic and semantic memory. If input were cued with either superordinate or coordinate cues, the difference between recall with the two types of cues found by Anderson with no cues at input should disappear or be reduced, since cues at input make storage of both types of cues in episodic memory more likely.

EXPERIMENT I

The present experiments used a more conventional cuing procedure than that used by Anderson. Lists of words were presented to be remembered. In Experiment I, cues were either presented or not presented during input. Output was cued with terms that were either superordinates or coordinates of the words to be remembered. Cues were always consistent at input and output, so that if superordinate cues were presented at input, superordinate cues were also presented at output. In order to determine whether the coordinate and superordinate cues were equally associated in semantic memory to the words to be remembered, subjects who had not seen those words were given the cues and asked to guess the words.

Method

Two lists of 20 words each were selected. For each word, both a superordinate and a coordinate cue were chosen. The words to be remembered had been used earlier in a task in which subjects were instructed to name a category for each of them. The categories named for each word were then rated by other subjects. In order for a term to be used as a superordinate cue in the present study, at least 10 of 100 subjects had produced the superordinate and the rating was at least 3.6 on a 5-point scale where 5 indicates 'a very good category name.' The mean production frequency of the superordinates selected was 42 out of 100 and the mean rating of the superordinates was 4.23. Coordinate cues had either never been named as categories for the words to be remembered or, if they had, were rated lower than 2 on a 5-point scale where 1 represents 'not a category name.' An attempt was made to control for the associability between the cues and words to be remembered by using superordinate and coordinate cues that were produced equally often as free associates to the words to be remembered.

In the conditions with cued input, cues were typed in lowercase letters above the words to be remembered, which were typed in uppercase letters. Stimuli were placed in slide mounts and presented at a 2-sec rate via a Kodak Carousel slide projector. All subjects were given instructions for free recall of the words to be remembered, and subjects in the cued conditions were told to use the terms typed in lowercase letters to help them remember the words in capital letters. After presentation of the words to be remembered, lists of the superordinate or coordinate cues were given to each subject, and subjects were instructed to use the cues to help them remember the words. Two minutes were allowed for written recall. There were two lists of words, two presentation orders, and two cue orders, so 8 subjects were necessary for balancing. Each of the four conditions consisted of 16 volunteers from introductory psychology classes at North Dakota State University; they participated for bonus points toward their grade.

After recall, each subject was given the superordinate and coordinate cues for the other list and asked to guess what words the other subjects had learned. Subjects in the superordinate conditions were given the superordinate cues for the other list and subjects in the coordinate condition were given the coordinate cues for the other list.

Results

The proportions of the words to be remembered that were correctly recalled and guessed in the four conditions are shown in Table 1. A word was counted as correct whether or not it was paired with its cue, although use of a stringent scoring by which a word had to be beside its cue did not alter the conclusions. An analysis of variance on the data on recall indicated that cues at input resulted in higher recall than no cues [$F(1, 60) = 12.56$, $p < .01$] and that superordinate cues produced higher recall than coordinate cues [$F(1, 60) = 16.82$, $p < .01$]. There was no interaction between these variables [$F < 1$]. While it appears that superordinates were better cues than coordinates, the data on guessing make any conclusions questionable. An analysis of variance on these data indicated that subjects guessed more words with superordinate than with coordinate cues [$F(1, 60) = 16.22$, $p < .01$].

Table 1. Proportion correct as a function of type of cue and presence of cues at input; Experiment I

Type of cue	Recall: Input cued	Recall: Input uncued	Guessing: Input cued	Guessing: Input uncued
Output cued				
Superordinate	69	60	.18	20
Coordinate	58	43	12	13

The results on recall basically replicated those of Anderson (1972), but his interpretation in terms of the superordinate relationship of a cue to a word is doubtful. The *associability*, not the superordinate relationship of cue to word, might better explain the differences in recall that were observed. Experiment II controlled for the associability of cues and words to see whether the superordinate or coordinate relationship per se plays any role in recall.

EXPERIMENT II

Each potential cue for a word to be remembered was presented to 100 subjects as a stimulus in a free-association task. One free associate was given by each subject to each stimulus. Superordinate and coordinate stimuli that produced the words to be remembered about equally often were selected for use as cues in the experiment. If the relationship between a cue and a word was critical to Anderson's results and the results of Experiment I, then superordinate cues should produce better recall than coordinate cues even with the associability between the cues and words equated.

Experiment II differed from Experiment I in two other ways: *mixed* lists were used, and cues were either present or absent at *output*. Anderson (1972) tested each word with each cue in a within-subjects design, and he found that superordinate cues were more effective than coordinate cues in this mixed list. To make Experiment II more similar to Anderson's study, half of the words in a list had superordinate cues and the other half had coordinate cues, so there was no consistent relationship between cues and words in a list. Experiment I was also expanded to include the presence and absence of cues at output as well as at input. Some studies have shown that cues at input but not at output actually hurt recall (e.g., Freund and Underwood, 1970), and Experiment II was designed to see whether or not the amount of such a decrement depends on the superordinate or coordinate relationship of the cues and the words.

Method

Two lists of 20 words to be remembered were selected from the free-association data described above. The criteria used in Experiment I to ensure that the relationship of cue to word was either superordinate or coordinate were used again. In each list, the 20 words were divided into two sets. Across subjects each set was used equally often with coordinate and superordinate cues. In the condition uncued at either input or recall, type of cue was a dummy variable but each subject's recall was scored as if one set of words

had been cued by coordinates and the other set cued by superordinates. One of two presentation orders was randomly assigned to each block of subjects in the four cuing conditions. Two different cue orders were used equally often in each condition. The procedure was identical to that used in Experiment I except that half of the subjects received no cues during output. After recall, subjects guessed words from the other list. Subjects were tested in pairs. There were 24 volunteers from introductory psychology classes in each of the four cuing conditions.

Results

The proportions recalled and guessed are shown in Table 2 for each condition. The only significant effect for recall was that of cues at output [$F(1, 92) = 18.19, p < .01$]. Cued output was better than uncued. There was no effect of type of cue [$F < 1$] and no decrement with the presence of cues at input but not at output. Neither the main effect of cues at input nor the interaction of cues at input by cues at output was significant [$Fs(1, 92) = 2.21$ and 3.51, $ps > .05$, respectively]. As can be seen in Table 2, the probabilities of guessing the words were quite low and unsystematic with respect to condition. An analysis of variance yielded no significant effects [all $Fs(1, 92) \leq 1.21$]. Hence, the degree of associability of cues and words was well equated for superordinate and coordinate cues. It is also clear that the superiority of superordinate over coordinate cues disappeared in this experiment.

DISCUSSION

The results of Experiment II offered no evidence for Anderson's conclusion that it is the relationship of cues and words to be remembered that is critical in recall. It may be that superordinate cues are generally

Table 2. Proportion correct as a function of type of cue and presence of cues at input and output (cuing conditions); Experiment II

	Recall		Guessing	
Type of cue	Input cued	Input uncued	Input cued	Input uncued
Output cued				
Superordinate	.57	.50	.04	.08
Coordinate	.64	.49	.06	.09
Output uncued				
Superordinate	.40	.42	.06	.07
Coordinate	.42	.42	.07	.06

better than coordinate cues, as was found in Experiment I and by Anderson (1972), because they tend to be more highly associated to the words than are coordinate cues. When this degree of association is controlled, however, the advantage of superordinate cues is lost.

Experiment II also differed from Experiment I in another way, which may have been a factor in removing the superiority of superordinate cues. In Experiment I, all words in a list were cued with either superordinate or coordinate cues. The relationship of cues and words may be more obvious with superordinate than coordinate cues, in that subjects can quickly identify the superordinate cues as the categories naming the words but find the coordinate cues more ambiguous, as perhaps similar or somehow related to the words. The more precise rule with superordinate cues may allow subjects to restrict their search to items that are instances of the category named by the cue. In Experiment II, on the other hand, no one consistent rule could be used to characterize the cues and the words, since the list was mixed. However, since Anderson tested each word with each cue in a within-subjects design and still found higher recall with superordinate than with coordinate cues, this explanation for the absence of a superiority of superordinate cues in Experiment II seems less likely than the explanation in terms of degree of association.

The view that it is the associability rather than the superordinate or coordinate relationship of the cue and the word to be remembered that is important in determining recall is not inconsistent with current theories of semantic memory. Collins and Loftus' (1975) model allows for links of different strengths and is not a rigid hierarchical model. Activation could begin from a node representing either a superordinate or a coordinate cue and proceed directly to the word to be remembered. The probability that the search would reach that word would depend on the strength of the link between it and the cue. This strength is roughly tapped by the free-association probability. When the strengths are about the same, superordinate and coordinate cues are about equally effective. As an explanation of the present data, this model of semantic memory comes very close to a generation/recognition model of cuing like that of Anderson and Bower (1972) or Bahrick (1970), whereby a subject would generate associates to a cue and then search among these associates until he recognizes the word. The more highly associated the cue and the word, the more likely the word would be among the alternatives generated, and hence, the more likely it would be recalled. With neither type of model is there a need to postulate the type of search processes described by Anderson, processes which predict the superiority of superordinate over coordinate cues.

This sort of an interpretation could easily be modified to fit Tulving's (1972) description of episodic memory. In the absence of cues at input, the probability of a cue being entered into episodic memory with a word depends on the strength of association of the cue and the word. In Experiment I, superordinate cues were more strongly associated to the words than were coordinate cues, and cued recall was higher with superordinate cues. The prediction that the difference between superordinate and coordinate cues should be larger with no cues at input than with cues at input was not supported, since there was no evidence for an interaction of type of cue by cues at input. However, the difference in degree of association between the two types of cues and their words was relatively small, as indicated by the data on guessing. Perhaps this difference in preexperimental associability was enough to boost recall with superordinate cues but was not large enough to produce the expected interaction. In Experiment II, no differences were observed between coordinate and superordinate cues either in recall or in guessing. With associability between the two types of cues and their words equated, they were equally likely to be entered in episodic memory and were equally effective as retrieval cues.

Regardless of which of these types of models is ultimately the most successful in explaining the growing data on cuing, the associability of cues and words to be remembered would seem sufficient to explain the effectiveness of various cues for a given word. Reference to the superordinate or coordinate relationship between cues and words would seem unnecessary.

Notes

A shortened version of this paper was presented at the meetings of the Midwestern Psychological Association in Chicago in 1976. Appreciation is expressed to William S. Maki, Jr., for his helpful comments on the manuscript. Requests for offprints should be addressed to the author, Department of Psychology, North Dakota State University, Fargo, North Dakota 58102. Received for publication June 8, 1976.

References

Anderson, J. R., and Bower, G. H. 1972. Recognition and retrieval processes in free recall. *Psychological Review* 79:97–123.

Anderson, R. C. 1972. Semantic organization and retrieval of information from sentences. *Journal of Verbal Learning and Verbal Behavior* 11:794–800.

Bahrick, H. P. 1970. Two-phase model for prompted recall. *Psychological Review* 77:215–222.

Collins, A. M., and Loftus, E. F. 1975. A spreading-activation theory of semantic processing. *Psychological Review* 82:407–428.

Freund, J. S., and Underwood, B. J. 1970. Restricted associates as cues in free recall. *Journal of Verbal Learning and Verbal Behavior* 9:136–141.

Quillian, M. R. 1966. Semantic memory. Doctoral dissertation, Carnegie Institute of Technology. Reprinted in part in *Semantic information processing,* ed. M. Minsky (Cambridge, Mass.: M.I.T. Press, 1968).

Thomson, D. M., and Tulving, E. 1970. Associative encoding and retrieval: Weak and strong cues. *Journal of Experimental Psychology* 86:255–262.

Tulving, E. 1972. Episodic and semantic memory. In *Organization of memory,* ed. E. Tulving and W. Donaldson. New York: Academic Press.

Tulving, E., and Osler, S. 1968. Effectiveness of retrieval cues in memory for words. *Journal of Experimental Psychology* 77:593–601.

AMERICAN JOURNAL OF PSYCHOLOGY
June 1977, Vol. 90, No. 2, pp. 291–298

Goalbox placements with initial nonreward and situational similarity in resistance to extinction

Joseph J. Franchina, James R. Weeks, Jr., and Nanette Pais
Virginia Polytechnic Institute and State University

Rats were given nonrewarded goalbox placements or nonrewarded instrument trials prior to continuously rewarded training and then underwent extinctic During initial nonreward, the external environmental stimuli were similar dissimilar to those of later extinction. Both types of initial nonreward reliat increased resistance to extinction relative to continuous reward alone and yield comparable levels of performance when the situation for initial nonreward w similar to that for later extinction. However, the goalbox placements yield reliably greater resistance to extinction in the dissimilar situation. Situatior similarity enhanced the effect of initial nonreward, but reliably so only with i strumental responding.

The administering of nonrewarded instrumental trials before continuous rewarded training has been shown to reliably increase resistance to extin tion as compared with continuously rewarded trials alone (Spear, Hi and O'Sullivan, 1965; Spear and Spitzner, 1967; Robbins, Chait, ar Weinstock, 1968; Capaldi and Haggbloom, 1974). According to Spe and Spitzner (1967), pretraining with nonreward provides exposure situational stimuli later present during extinction and thereby serves to r duce the decremental effects of the shift from continuously reward training to nonrewarded extinction. If this effect of initial nonreward fostered by the comparability of the situation for pretraining and the sit ation for extinction, then Spear and Spitzner's proposal suggests that t greater the similarity between the two situations, the greater the resistan to extinction. One purpose of the present experiment was to evaluate tl possibility.

In prior studies of the effect of initial nonreward, the nonreward event has typically been presented after the performance of some instr mental response, like running in a runway. This procedure makes somewhat difficult to ascertain whether the effect is due to the tande occurrence of instrumental responding *and* nonreward or to the occu rence of nonreward per se. Conceivably, like other extinction effects, t effect of initial nonreward may be demonstrable after procedures th

preclude, or minimize, performance of the full instrumental response (for evidence on the effect of partial reinforcement on extinction, PREE, see Brown and Logan, 1965; McCain, Baerwaldt, and Brown, 1969; and Capaldi, 1971). To evaluate this possibility, the present experiment investigated the effect of administering initial nonreward by placing the animal directly in the goalbox, thereby precluding its performance of the full instrumental response.

METHOD

Apparatus

Two alleyways were used. One, alley A, was a modified version of that described by Franchina and Brown (1970). It was constructed of wood, painted flat black, and measured 146.0 cm long by 9.0 cm wide by 10.0 cm high. Guillotine doors divided the alley into a 32.0-cm startbox; a 90.9-cm runway; and a 23.1-cm goalbox, attached to the runway at a 90-deg angle to the left. The alley floor was stainless-steel rods. In the goalbox, a Plexiglas barrier, 4.5 cm high by 9.0 cm wide by .16 cm thick, was placed 6.4 cm in front of the rear wall of the box.

The other alleyway, alley B, was constructed of Plexiglas, painted with black and white vertical stripes about 1.9 cm wide. Alley B was 146.5 cm long and was divided by guillotine doors into a startbox, 32.1 cm by 12.8 cm by 10.9 cm; a runway, 86.0 cm by 9.0 cm by 10.9 cm; and a goalbox, 27.9 cm by 12.8 cm by 10.9 cm. Each section of alley B, and of alley A, was covered with clear Plexiglas tops. The floors of the startbox and goalbox of alley B were Masonite, painted white; the floor of the runway was hardware cloth, painted black. In the goalbox, a Plexiglas barrier, 3.8 cm high by 12.8 cm wide by .16 cm thick, was placed 6.4 cm in front of the rear wall of the box.

In each alley, locomotor performance was measured by two photocells in circuit with a timer, calibrated in .01-sec units. One photocell was located 33.0 cm from the rear wall of the startbox; the other photocell was 102.8 cm from the first and 3.8 cm in front of the barrier in the goalbox. Interruption of the first photobeam started the timer; interruption of the second beam stopped the timer.

Subjects

Sixty experimentally naive male hooded rats, 90–100 days old, were obtained from the local departmental colony. Each rat was housed individually and was fed a daily ration of wet mash, 10 g of dry ground Chow mixed with two tablespoons of water. Daily feedings occurred 15–30 min after experimental treatments. Water was always available ad lib.

Design and procedure

All rats received five days of habituation to the feeding regimen, including three days of exposure in the home cage to the reward to be used later in training (20 .045-g Noyes pellets). Then, on the sixth day, the rats were randomly assigned to experimental groups and were started on phase 1 of a three-phase experiment. There were four initially nonrewarded groups and two continuously rewarded control groups, 10 rats per group.

Phase 1

For five days, two experimental groups received four *nonrewarded instrumental* trials in the runway (condition IN); two other experimental groups received four *nonrewarded placements* in the goalbox (condition PN) under procedures that precluded performance of the full instrumental response of traversing the entire alley. Half of the rats under conditions IN and PN, 10 from each, received nonreward in a situation highly *similar* to that of later nonrewarded extinction (condition S); the remaining rats under conditions PN and IN received nonreward in a situation highly *dissimilar* to that of later extinction (condition D). For the controls, one group, C40, received four continuously rewarded instrumental trials on each day in phase 1 (the reward was 20 pellets per trial); the other group, C20, received only handling on each day. For the groups under conditions PN and IN, alleys were counterbalanced across conditions S and D. For group C40, and later for group C20, half of the rats were trained and extinguished in alley A throughout the experiment; the other half, in alley B. Analysis of variance within appropriate experimental and control groups showed that alley per se had negligible effects on performance [$Fs < 1$].

For a nonrewarded goalbox placement, the rat was removed from the home cage; placed in the appropriate goalbox, its head down directly over an empty glass caster; and confined for 20 sec. For a nonrewarded instrumental trial, the rat was placed into the startbox of the appropriate alley; the guillotine door was raised, and the rat was allowed 60 sec to interrupt the photobeam in the goalbox. If it did, the goalbox door was closed immediately, and the rat was confined for 20 sec with an empty glass caster. For the instrumental trials, the procedures essentially followed those described by Franchina and Brown (1970). For group C40, these trials were like the nonrewarded instrumental trials except for the reward. For all groups, the interval between goalbox placements or between instrumental trials was approximately 12 min, spent in the home cage.

Phase 2

Phase 2 started on the day after phase 1 ended. In phase 2, all groups received five days of four continuously rewarded training trials per day; the reward was 20 pellets per trial. These procedures followed those described for group C40 in phase 1. In phase 2, each rat was trained in the same alley, A or B, as that to be used in extinction.

Phase 3

On the day after phase 2 ended, phase 3 began. All groups received three days of five extinction trials per day according to the procedures described in phase 1 for nonrewarded instrumental trials. The measure of performance was total running time. Transformations of the latency data into logarithms yielded essentially the same results as the raw data presented below.

RESULTS

Phase 1

The performance of groups INS and IND on the initially nonrewarded instrumental trials was similar and showed negligible change over blocks

of trials [$Fs < 1$]. The latter finding is consistent with earlier data from Capaldi and Haggbloom (1974), Spear et al. (1965), and Robbins et al. (1968). Mean total running times for groups INS and IND combined were 29.0, 38.9, 35.2, 34.1, and 34.9 sec for blocks 1–5. The performance of group C40 was similar to that shown by group C20 in phase 2. Mean total running times for group C40 on blocks 1–5 of phase 1 were 17.5, 18.2, 8.6, 6.9, and 5.0 sec.

Phase 2

Figure 1 presents performance in phase 2 and in phase 3 (extinction) for the initially nonrewarded groups and the controls. The initially nonrewarded groups (INS, PNS, IND, and PND) performed similarly to each other in phase 2. Analysis of variance revealed no reliable differences

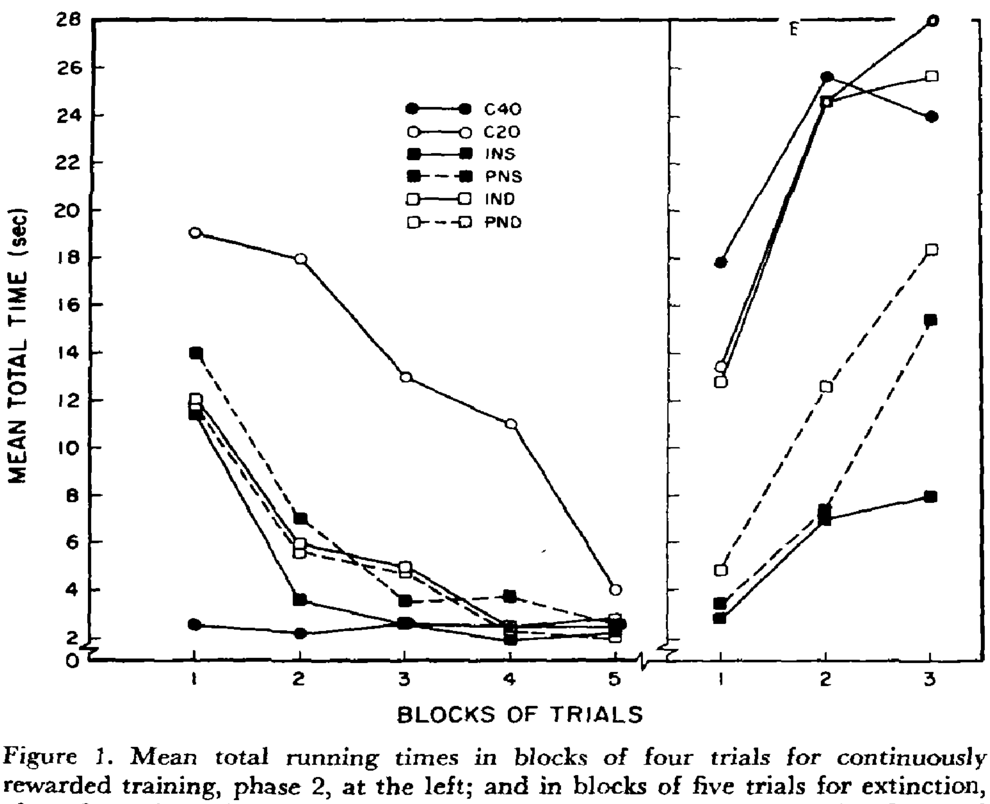


Figure 1. *Mean total running times in blocks of four trials for continuously rewarded training, phase 2, at the left; and in blocks of five trials for extinction, phase 3, at the right:* data shown for continuously rewarded controls (C40 and C20) and for initially nonrewarded groups (N) with placements in the goalbox (P) or instrumental trials in the runway (I) in a situation similar (S) or dissimilar (D) to that of extinction

among these groups due to treatment in phase 1 [$Fs < 1$]. The results of phase 2 are pertinent, however, to verifying previously reported results of initial nonreward (e.g., Spear et al., 1965; Spear and Spitzner, 1967). Briefly, group INS ran more slowly than group C40 on blocks 1 and 2, and more rapidly than group C20 on blocks 1, 2, 3, and 4. Analysis of variance and simple-effects comparisons showed that the difference between groups INS and C20 was reliable [$p < .025$] on each block of trials. This finding suggests that initially nonrewarded instrumental trials facilitate rather than impair subsequent responding under continuous reward (Spear and Spitzner, 1967). Trial-by-trial comparisons between groups INS and C40 further showed that on the first trial of blocks 1, 2, and 3, group INS ran more slowly than group C40; but by the fourth trial of the same blocks, groups INS and C40 performed indistinguishably from each other. Apparently, the first-trial decrement of group INS dissipated rapidly over subsequent trials (Spear et al., 1965). Statistical comparisons showed that the difference between groups INS and C40 was reliable on the first trial [$p < .01$], but not on the fourth trial [$p > .05$, at least], of blocks 1, 2, and 3.

Phase 3

A comparison of the data in Figure 1 for groups INS, C40, and C20 shows the effect of initial nonreward on performance during extinction. Nonrewarded instrumental trials before continuously rewarded training (group INS) increased resistance to extinction relative to controls with continuously rewarded training only (groups C40 and C20). The inclusion of group C20 in this comparison verifies that the effect was not due simply to a difference between groups INS and C40 in total number of continuously rewarded trials but, rather, probably reflects the influence of pretraining under nonreward. Groups INS and C20 received the same number of continuously rewarded trials. Analysis of variance over all the extinction data of groups INS, C40, and C20 yielded a reliable effect of groups [$F(2, 27) = 74.26$, $p < .001$]. In simple-effects comparisons, group INS differed reliably from groups C40 and C20 [$p < .01$], but groups C40 and C20 did not differ reliably from each other [$p > .20$].

For the initially nonrewarded groups, Figure 1 shows that goalbox placements and instrumental trials (groups PNS and INS) produced roughly equivalent levels of extinction when the situations for initial nonreward and extinction were similar to each other. However, when these situations were dissimilar, initially nonrewarded goalbox placements (group PND) produced better performance during extinction than did nonrewarded instrumental trials (group IND). In general, then, the

greater the similarity between the situations for initial nonreward and later extinction, the better the performance in extinction. However, this effect was larger for the instrumental trials than for the goalbox placements.

Analysis of variance over all the data on extinction of the initially nonrewarded groups showed reliable effects for the interactions of initial nonreward (P or I) by similarity (S or D) and blocks by initial nonreward by similarity [$Fs(1, 36; 2, 72) = 12.67$ and 9.58, $ps < .001$]. Simple-effects comparisons showed that groups PND and IND differed reliably from each other [$F = 10.27$, $p < .001$], but groups PNS and INS did not [$F = 2.50$, $p > .10$]. Group INS differed reliably from group IND [$F = 37.42$, $p < .001$], but groups PNS and PND did not differ reliably from each other [$F = 2.08$, $p > .10$]. The degrees of freedom for each of the simple-effects comparisons were 1 and 18. Each comparison also yielded a reliable effect of blocks, but the interaction of blocks by groups was not reliable in any comparison. Only the comparison of groups PNS and PND even neared reliability [$F(2, 36) = 3.00$, $p = .05$–$.07$].

Finally, the unexpected superiority of group PND over group IND prompted a further evaluation in terms of individual subjects. Specifically, the performance of each rat in group PND was compared against that of the fastest rat in group IND. Mean total times for the fastest rat in group IND were 7.3, 14.4 and 10.0 sec on blocks 1, 2, and 3 during extinction. Considering performance over all three blocks of trials, 7 of 10 rats in group PND ran faster than the fastest rat in group IND. This finding indicates that the performance of individual subjects tended to substantiate the difference between groups PND and IND shown in the group data of Figure 1.

DISCUSSION

This study showed that the greater the similarity between external environmental stimuli during initial nonreward and later extinction, the more initial nonreward enhanced performance during extinction. While this enhancement was reliable only for initially nonrewarded instrumental trials, the goalbox placements and instrumental trials both yielded results consistent with Spear and Spitzner's (1967) explanation of the effect of initial nonreward. That is, the nonreward presumably provides familiarity with stimuli present later in extinction and thereby curtails the decremental effects expected following the change in stimuli from continuously rewarded training to nonrewarded extinction. In this study, nonreward was administered to groups INS, PNS, IND, and PND; but groups INS

and PNS received the nonreward in the presence of environmental stimuli (e.g., visual cues about the apparatus) that were highly similar to those of later extinction. Consequently, the experience of initial nonreward by groups INS and PNS may have generalized more readily to extinction to facilitate their performance during extinction over that of groups IND and PND, both of which received initial nonreward in the presence of environmental stimuli that were dissimilar to those of extinction.

In this study, the effect of initial nonreward was produced by nonrewarded goalbox placements in pretraining. These results are consistent with reports (e.g., McCain et al., 1969) of the results of such placements on the effect of partial reinforcement on extinction. Apparently, neither effect necessarily depends on performance of the full instrumental response. On the other hand, it would be gratuitous to disclaim any role for instrumental performance in producing the effect of initial nonreward. The goalbox placements in this study precluded the full instrumental response of traversing the entire alley, but they did not prevent locomotor behavior *in* the goalbox. Capaldi (1971), Theios and Polson (1962), and Trapold and Doren (1966) have indicated that the effect of such placements on extinction is influenced by the extent to which locomotor behavior occurs in the goalbox. Accordingly, in the present study the amount of locomotor behavior in the goalbox by groups PNS and PND in pretraining may have been sufficient for the generalization of responding to later extinction. Thus, the goalbox placements enhanced responding during extinction.

Further, if activity in the goalbox does influence the effect of initial nonreward and traversing the entire alley is inessential, then the comparable levels of performance during extinction by groups PNS and INS should not be surprising. Although group INS was allowed to traverse the alley, it was then confined, without reward, to the goalbox for 20 sec just as was group PNS. The critical variable, goalbox experience, was the same for both groups.

The most puzzling finding of this experiment was that with dissimilar environments during initial nonreward and extinction, the nonrewarded goalbox placements yielded reliably *better* performance during extinction than the nonrewarded instrumental trials. This finding seems contrary to the proposition that an influential factor in the effect of initial nonreward is activity in the goalbox. According to this proposition, groups PND and IND should show the same general relationship in extinction as that between groups PNS and INS. One explanation for the difference between groups PND and IND might be that even if activity in the goalbox was similar for groups PND and IND, group IND's additional exposure to environmental cues in the runway may have increased the dis-

criminability of the initial nonreward situation from that of later extinction, whereas group PND's exposure was limited to the cues in the goalbox. Thus, greater exposure to more cues of the external environments for group IND than for group PND may have increased the dissimilarity between initial nonreward and later extinction situations for group IND and consequently curtailed the generalization of initial nonreward relative to that for group PND.

Notes

Requests for offprints should be addressed to J. J. Franchina, Department of Psychology, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061. The data were reported at the Southeastern Psychological Association Convention in Atlanta in April, 1975. The authors thank W. B. Pavlik for critically reading this manuscript. Received for publication January 11, 1976; revision, August 29, 1976.

References

Brown, R. T., and Logan, F. A. 1965. Generalized partial reinforcement effect. *Journal of Comparative and Physiological Psychology* 60:64–69.

Capaldi, E. D. 1971. Effect of nonrewarded and partially rewarded placements on resistance to extinction in the rat. *Journal of Comparative and Physiological Psychology* 76:483–490.

Capaldi, E. J., and Haggbloom, S. J. 1974. Influence of reward magnitude on the initial nonreward effect. *Bulletin of the Psychonomic Society* 3:219–221.

Franchina, J. J., and Brown, T. S. 1970. Response patterning and extinction in rats with hippocampal lesions. *Journal of Comparative and Physiological Psychology* 70:66–72.

McCain, G., Baerwaldt, J., and Brown, E. R. 1969. Extinction following a small number of goal-box placements. *Canadian Journal of Psychology* 23:274–284.

Robbins, D., Chait, H., and Weinstock, S. 1968. Effects of nonreinforcement on running behavior during acquisition, extinction, and reacquisition. *Journal of Comparative and Physiological Psychology* 66:689–706.

Spear, N. E., Hill, W. F., and O'Sullivan, D. J. 1965. Acquisition and extinction after initial trials without reward. *Journal of Experimental Psychology* 69:25–29.

Spear, N. E., and Spitzner, J. H. 1967. Effects of initial nonrewarded trials: Factors responsible for increased resistance to extinction. *Journal of Experimental Psychology* 74:525–537.

Theios, J., and Polson, P. 1962. Instrumental and goal responses in nonresponse partial reinforcement. *Journal of Comparative and Physiological Psychology* 55:987–991.

Trapold, M. A., and Doren, D. G. 1966. Effect of noncontingent partial reinforcement on the resistance to extinction of a runway response. *Journal of Experimental Psychology* 71:429–431.

AMERICAN JOURNAL OF PSYCHOLOGY
June 1977, Vol. 90, No. 2, pp. 299–308

Experimental control of reaction times to negative and expletive sentences

Ryan D. Tweney and Dan Swart
Bowling Green State University

In a study of the effects of instructions on reaction times to judgments of the truth or falsity of sentences, 40 undergraduates were provided computer-assisted instruction by either the 'true' or 'conversion' model and required to judge 64 sentences of all possible combinations of true or false, affirmative or negative, and expletive or nonexpletive. The pattern of reaction times corresponded exactly to that predicted by the type of instructions, including the reaction times to expletive sentences, which extends the range of syntactic structures to which the models apply. Such experimental control has implications for a broad range of methodological problems in cognitive psychology.

Measurement of the time between presentation of a sentence and a judgment of the sentence's truth or falsity has been a useful technique in the investigation of human cognition. Comparison of reaction times to sentences of differing kinds has allowed inferences about the stages that are involved in processing the information contained within the sentences and in responding appropriately to that information. Such inferences rest on the validity of two assumptions: that the stages in processing are independent of one another and that processing occurs serially in real time (Sternberg, 1969, 1975).

The accurate prediction of reaction times to negative and positive sentences has been repeatedly demonstrated (Carpenter and Just, 1975; Clark, 1974; Wason and Johnson-Laird, 1972). Consider a sentence like 'Three is not an odd number.' To successfully judge this as a false sentence, a subject must integrate the information that three is indeed an odd number with the information that the sentence is negating this fact. It is thus reasonable to expect that such a sentence will take longer to process than a positive sentence like 'Three is an odd number,' and this result has been observed (Trabasso, 1972, reviews a number of studies). However, it is also the case that a judgment of falsity (e.g., 'Three is even') takes longer than a judgment of truth (e.g., 'Three is odd'). The

question then arises as to how judgment of the truth or falsity of a sentence is related to the processing of a negative term like 'not.' Two models of the relationship have been elaborated by Clark (1974).

According to the *true* model, all sentences are first encoded as a central proposition, together with an embedding structure containing any negative elements. The truth or falsity of the central string is determined and an appropriate response chosen. If the embedding string contains a negative, the chosen response is changed to its opposite. Thus, 'Three is not even' is encoded as 'Not [three is even],' 'false' is chosen as the response to the central string, and is then changed to 'true' because of the presence of 'not' in the embedding string. According to an alternative model, the *conversion* model, all sentences are encoded by first converting any negative predicates to their positive counterparts and then determining a response on the basis of the resultant sentence. Thus, 'Three is not even' is converted to 'Three is odd,' which is then evaluated as 'true.'

The two models make different predictions about reaction times to certain sentences. Both models predict that reaction times to true affirmative (TA) sentences like 'Three is odd' will be faster than those to false affirmative (FA) sentences like 'Three is even,' since falsity takes longer to process by both models. Likewise, both models predict that negative (TN or FN) sentences will take longer than affirmative (TA or FA) sentences, since one extra step is involved. The conversion model, however, predicts that reaction times to true negatives (TN) such as 'Three is not even' will be faster than reaction times to false negatives (FN) such as 'Three is not odd,' since the resulting sentence after conversion of a false negative is itself false. In effect, false negatives require two steps (conversion and processing a false statement), whereas true negatives require only one (conversion). The so-called true model predicts the reverse order of reaction times to these sentences. The embedded string of a true negative is false, requiring more time to process. Hence, true negatives should take *longer* than false negatives, since they involve two steps (processing a false statement and incorporating a negative element), whereas false negatives require only one (incorporating a negative element). In effect, the true model predicts an interaction of negative or affirmative by true or false, whereas the conversion model predicts additivity of effects.

Clark presented evidence confirming the predictions made by the true model. Trabasso (1972) summarized studies confirming both models, and he suggested that one or the other strategy may operate depending on the situational context. A similar, more general argument was made by Glucksberg, Trabasso, and Wald (1973), who suggested that the depth and type of information processing in sentence-verification tasks may depend both on the situational context and on the demands of the task.

Clark (1974) further claimed that the type of processing engaged in by the subject could be directly controlled by appropriate instructions. Such instructions were described as "quite possible to follow" (Chase and Clark, 1972, p. 211), though no quantitative data were presented and no report of the relevant study (by Young and Chase) has ever been published. While Clark (1974) indicated that 95% of the variance was accounted for in Young and Chase's experiment, it is not clear from the context whether the instructions alone were responsible. Finally, two pilot studies in our laboratory failed completely to achieve such control, suggesting a need for replication. If replicable, the finding is of special interest because it suggests that the kind of stages in processing described by each model may correspond to psychological units that are under voluntary control and representable in the conscious awareness of the subject. The present study therefore attempted replication by use of extensive training.

In addition to replication, the model was extended to another type of sentence than had previously been examined. Clark (1973) has commented on the necessity for such extension in psycholinguistic research to avoid 'fixed effect' fallacies — erroneous overgeneralizations based on effects that are in actuality limited to specific sentences or specific grammatical forms. Expletive sentences, of the form 'It is the case that three is odd' and 'It is not the case that three is even,' provide an interesting variant of sentences of the primary type (TA, FA, TN, and FN). Four variations of the expletive type exist (TAE, FAE, TNE, and FNE), and exactly the same predictions can be made about the rank order of reaction times. Since TNE and FNE sentences separate the negative element from the rest of the sentence, they provide an even stronger test of whether or not reaction times can be brought under experimental control. Furthermore, if the pattern of reaction times to expletive sentences differs from that for the primary sentences, then it would follow that aspects of the surface form of sentences are involved in the processing of negatives. Such a result would disconfirm both of the current models, since each model assumes that all relevant processing occurs only at the level of deep, or underlying, structure.

METHOD

Subjects

Subjects were 12 male and 28 female undergraduates enrolled in an introductory psychology course. All received course credit for participation in the experiment.

Apparatus

The sentences were generated by a Data General NOVA 1220 minicomputer and presented visually using an Owens-Illinois Digivue plasma display device controlled by the computer. Letters were formed from suitably arranged 6-by-9 arrays of dots and appeared bright orange against a dark background. The sentences were written on the screen from left to right, the time to write them ranging from approximately 20 msec to approximately 90 msec, depending on each sentence's length. The room's illumination was kept at a constant low level to heighten the display's contrast. Subjects responded by pressing a key on the Digivue keyboard. Reaction time was measured by an internal 100-Hz clock in the computer. The clock started when the sentence was presented, and it stopped when the subject pressed a key. Measurements were accurate to the nearest .01 sec.

Stimulus materials

The stimuli consisted of the four sentences of the primary type, TA, FA, TN, FN, and their expletive counterparts, TAE, FAE, TNE, and FNE. Only the digit 1 was used in the sentences during training. The digits 2–9 were used in the experimental sentences. Each subject received all 64 possible sentences.

Procedure

The 40 subjects were randomly assigned to one of two training groups of 20 each (6 males and 14 females). Presentation of the sentences was randomized separately for each subject, using a programmed random-number generator. Each sentence remained visible until the subject pressed a key. Subsequent sentences appeared approximately 1.5 sec after each response. The '1' and '0' keys on the display keyboard were used to indicate 'true' and 'false' as responses. These response keys were alternated such that half of the subjects in each group indicated 'true' with the dominant hand and half with the nondominant hand. Computer software was controlled by the experimenter, using a teletypewriter positioned approximately 3 ft to the right of the subject.

Each set of instructions for training provided practice on each of the stages in processing called for by the model. Subjects were shown examples of each stage, required to practice each using an overt response, and finally, required to internalize each stage by practicing covert responses. As each stage was mastered it was combined with previously learned stages. Instructions for each experimental group were matched for examples presented, for total responses required of each group at each stage of training, and for overall length. All examples used only the digit 1. Each subject was presented training with the stated rationale of ensuring correct, quick responding.

Training by the true model consisted of three stages. (1) Evaluation of the sentence as though 'not' were absent. Subjects first responded out loud ('true' or 'false') to eight examples, then were asked to respond silently to themselves to eight further examples. (2) Evaluation of the sentence, as before, followed by separate evaluation of the sentence with 'not' present. Subjects first responded overtly to eight sentences, then covertly to eight more. (3) Evaluation, as before, in two steps, but with a key press to indicate truth or falsity. Sixteen examples were presented.

Training by the conversion model also consisted of three stages. (1) Conversion of the sentence followed by vocalization of the result of conversion ('One is odd' or 'One is even'). After eight overt examples, covert responding was requested for an additional eight. (2) Conversion, as before, followed by vocal evaluation of the results of the conversion ('true' or 'false'). Eight overt and eight covert examples were given. (3) Conversion and evaluation, followed by a key press to indicate truth or falsity. Sixteen examples were presented.

RESULTS

Analysis of error rates indicated that there were no significant differences in number of errors as a function of any of the conditions, ruling out the possibility of a trade-off of speed and accuracy in the present results. Consequently, only correct reaction times are included in subsequent analysis. Mean reaction times for the group of subjects trained by the true model are shown in Figure 1 and for the group trained by the conversion model in Figure 2. All predictions were confirmed: the rank orders predicted by the training procedure were manifested in the results of both

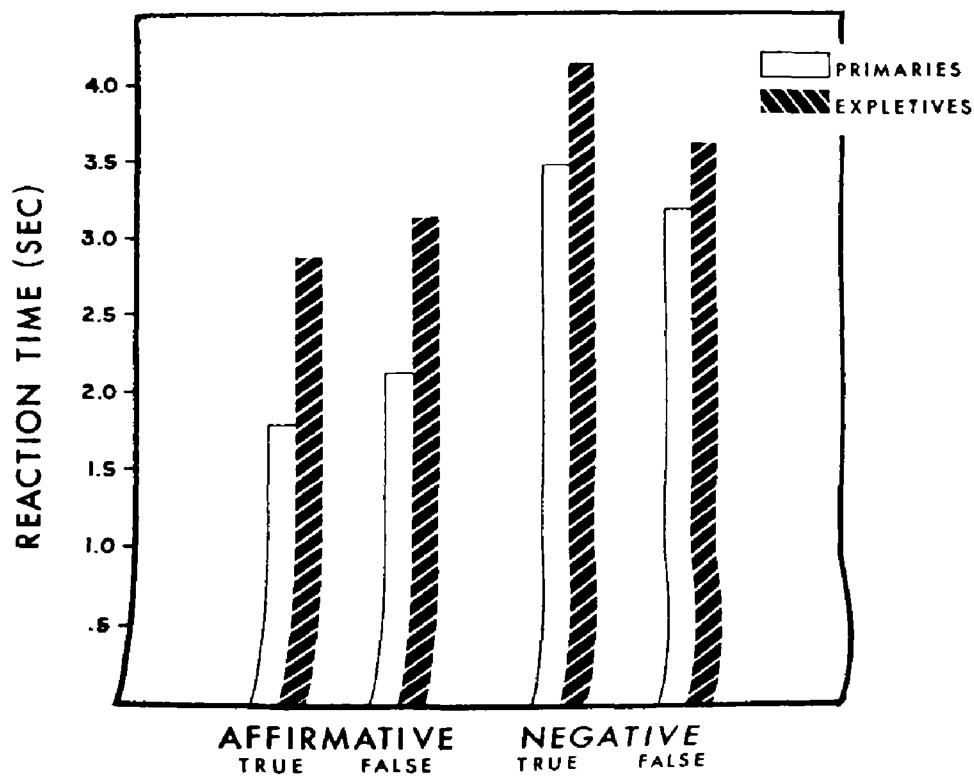


Figure 1. Mean reaction times for subjects given true-model training

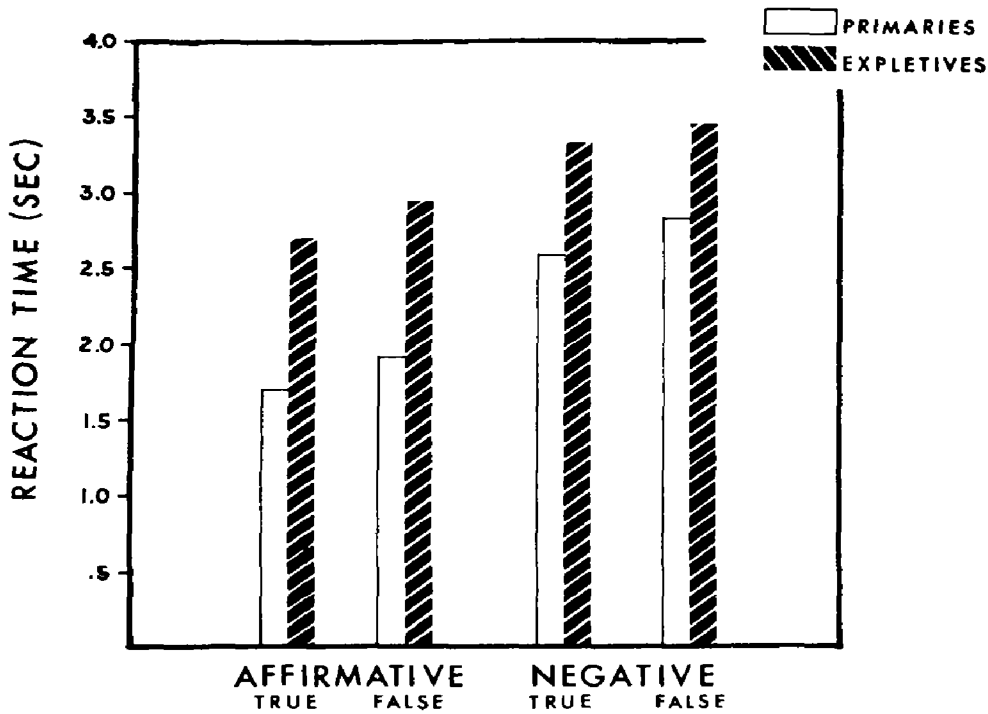


Figure 2. Mean reaction times for subjects given conversion-model training

groups of subjects and for sentences of both the primary and expletive types. Analysis of variance indicated significant main effects for training procedure [$F(1, 38) = 4.405, p < .05$], for sentence type [$F(1, 38) = 173.04, p < .001$], and for affirmative versus negative [$F(1, 38) = 141.71, p < .001$], but not for true versus false [$F(1, 38) = 3.70, p > .05$]. Training procedure interacted with affirmative versus negative [$F(1, 38) = 9.60, p < .005$] and with true versus false [$F(1, 38) = 9.25, p < .005$], but not with sentence type [$F(1, 38) = 1.21, p > .05$]. Sentence type interacted with affirmative versus negative [$F(1, 38) = 6.02, p < .05$], but not with true versus false. True versus false did, however, interact with affirmative versus negative [$F(1, 38) = 24.06, p < .001$]. Only one three-way interaction was significant, that of training procedure by true versus false by affirmative versus negative [$F(1, 38) = 7.78, p < .01$]. The four-way interaction was not significant.

Simple interaction effects were computed as direct tests of the principal hypothesis. The interaction of affirmative versus negative by true versus false was significant for those subjects trained by the true model [$F(1, 38) = 29.60, p < .001$], but not for those trained by the conversion model

[$F(1, 38) = 2.23$, $p > .05$]. This result is consistent with the hypothesized outcome: TN sentences were faster than FN sentences after training by the true model, but the reverse was true after training by the conversion model.

Do the results after training reflect the kind of processing characteristic of subjects who have *not* received training? To answer this question, a control group was run consisting of 10 male and 11 female subjects drawn from the same subject pool. The control subjects were given instructions that included all of the examples given to the experimental subjects and were approximately the same in overall length. However, no explicit instruction was provided in methods of solving the sentences. Subjects in the control group made many more errors (10.2%) than did subjects with training by either the conversion (3.7%) or true model (4.8%). For this reason, no analysis of variance was conducted. Control subjects took uniformly longer to respond to all eight types of sentences and manifested the pattern predicted by the true model (TN slower than FN) for both primary and expletive sentences.

DISCUSSION

It is evident from the results that experimental control of the pattern of reaction times is, as Chase and Clark stated, "quite possible." The repeated failures in our pilot studies suggest, however, that explicit training, in which the subject is separately instructed on each part of the algorithm for solution, may be necessary to achieve the result.

In addition, it is clear that both of the models generalized perfectly to expletive sentences. Expletives took consistently longer to evaluate than primary sentences, but the pattern of reaction times was identical. The interaction of sentence type by affirmative versus negative resulted from the somewhat longer times for affirmative expletives. Why should the difference between sentences like 'Three is even' (TA) and 'It is the case that three is even' (TAE) be greater than the difference between ones like 'Three is not even' (TN) and 'It is not the case that three is even' (TNE)? It is possible that subjects double-check affirmative expletives to see if any negative markers have been missed. This could account for the interaction, since such checking would not be needed for negative expletives. With this minor exception, the results confirm the usual assumption that the processing of negative sentences occurs only at the level of deep structures.

While the pattern of reaction times in the present study mirrors that found in other research, the quantitative relationships between conditions

are not typical. In particular, the latencies in our study (which averaged 2.84 sec) were longer than those typically observed, as indicated by the summary tables of previous findings provided by Carpenter and Just (1975, pp. 58–60, 67). It is likely that the discrepancy is a result of the explicit training provided our subjects, which may have had the effect of making them more cautious in responding.

Carpenter and Just (1975) argued that the difference in times required to judge negative and positive sentences true or false may reflect the different number of times that a single mental process, a retrieve-and-compare operation, is applied. While their model accurately predicted a number of the parametric differences between groups in their study and in others, it does not seem applicable to our data. The observed negation times (the average of TN − TA and FN − FA) and falsification times (the average of FA − TA and FN − TN) for our data are shown in Table 1. The negation times were generally longer than the falsification times, as expected, but the ratio of the two presents a puzzling set of relationships. Carpenter and Just found that such ratios were always either 2:1 or 4:1, reflecting the number of times that the retrieve-and-compare operation was applied. Yet for our data, none of the ratios seem to be fully consistent with this pattern. For this reason, it is unlikely that our subjects were utilizing a similar retrieve-and-compare operation. A uniprocess model does not, in short, appear to be applicable to our data.

The successful manipulation of the pattern of reaction times suggests a close correspondence between the conscious character of each information-processing stage and the possibility of bringing the stages under experimental control. The experimental instructions for the training in the present experiment first directed a subject's awareness toward the kind of processing needed at each stage, then provided practice utilizing overt responding, and finally provided covert, internalized practice. As each stage was correctly internalized by the subject, it was added to previously internalized stages. The process is reminiscent of Vygotsky's (1934) description of the internalization of verbal control. The result is of special interest because cognitive theories do not generally specify which stages

Table 1. Negation and falsification times (msec)

	True-model training		Conversion-model training	
	Primaries	Expletives	Primaries	Expletives
Negation times	1,380	906	582	760
Falsification times	320	367	191	166
Negation : falsification times	4.31:1	2.47:1	3.05:1	4.59:1

may or may not be represented in conscious awareness. There is, of course, good reason for this, given the notorious difficulties of introspective data. Direct controls of the kind used in this study may be a way to overcome this difficulty. To the extent that a particular manipulation of information can be taught a subject by explicit training procedures, then the particular stages must be consciously representable. This does not, of course, rule out the possibility that particular subparts of each stage go on below awareness.

It is interesting to note that the present result represents the kind of direct experimental control of behavior usually found only in more restricted operant situations. While the possibility of such control has sometimes been used as a criterion for the adequacy of 'understanding' of phenomena (Sidman, 1960), such control has not generally been sought over the stages in cognitive processing.

Evaluations of this sort should be especially valuable in those areas of cognitive psychology where differences in processing result from situational demands or from individual differences in processing. Glucksberg et al. (1973) provided examples of the effects of situational demands on processing. Individual differences in cognition have long been recognized (Galton, 1883), though little research has been directed to such questions in recent years. Yet, to the extent that such differences exist, it is inappropriate to attempt to control for their effect by averaging over groups of subjects. Not only does this increase the amount of 'error' variance, it leads to the possibility that the aggregate data will fail to reflect the characteristics of any particular subgroup of subjects. This, in effect, is exactly the problem encountered by the aggregate learning curve (Kimble, 1961). In cognitive research, this difficulty may be avoidable by direct controls of the kind described in this paper.

Notes

The authors thank Philip Lichtenfels for his assistance in gathering the data.
Offprints may be obtained from the first author, whose address is Department of Psychology, Bowling Green State University, Bowling Green, Ohio 43403.
Received for publication July 16, 1975; revision, July 29, 1976.

References

Carpenter, P. A., and Just, M. A. 1975. Sentence comprehension: A psycholinguistic processing model of verification. *Psychological Review* 82:45–73.

Chase, W. B., and Clark, H. H. 1972. Mental operations in the comparison of

sentences and pictures. In *Cognition in learning and memory,* ed. L. W. Gregg. New York: Wiley.

:lark, H. H. 1973. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior* 12:335–359.

:lark, H. H. 1974. Semantics and comprehension. In *Current trends in linguistics,* vol. 12, *Linguistics and adjacent arts and sciences,* ed. T. A. Sebeok. The Hague: Mouton.

ralton, F. 1907. *Inquiries into human faculty and its development.* London: J. M. Dent. First published in 1883.

:lucksberg, S., Trabasso, T. and Wald, J. 1973. Linguistic structures and mental operations. *Cognitive Psychology* 5:338–370.

:imble, G. A. 1961. *Hilgard and Marquis' "Conditioning and learning."* 2d ed. New York: Appleton-Century-Crofts.

idman, M. 1960. *Tactics of scientific research: Evaluating experimental data in psychology.* New York: Basic Books.

ternberg, S. 1969. Memory scanning–mental processes revealed by reaction-time experiments. *American Scientist* 57:421–457.

ternberg, S. 1975. Memory scanning: New findings and current controversies. *Quarterly Journal of Experimental Psychology* 27:1–32.

'rabasso, T. 1972. Mental operations in language comprehension. In *Language comprehension and the acquisition of knowledge,* ed. J. B. Carroll and R. O. Freedle. New York: Wiley.

/ygotsky, L. S. 1962. *Thought and language.* Cambridge, Mass.: M.I.T. Press. First published, in Russian, in 1934.

Vason, P. C., and Johnson-Laird, P. N. 1972. *Psychology of reasoning: Structure and content.* Cambridge, Mass.: Harvard University Press.

AMERICAN JOURNAL OF PSYCHOLOGY
June 1977, Vol. 90, No. 2, pp. 309-312

'Forced' encoding but 'no' learning

Allen R. Dobbs
University of Alberta, Canada

Two-dimensional stimuli consisting of a color and a form were paired with single digits as responses, but in an unconventional manner. For three of the items, color was the relevant component, and for the other three items, form was the relevant component. The other components were re-paired on successive trials such that they were orthogonal to correct responding. Only 2 of 20 subjects learned the experimenter-defined list in the allotted 40 trials. Instead of selecting bidimensionally (three relevant colors and three relevant forms), they selected components from a single dimension, thereby precluding their learning the list. The theoretical implications of this selection behavior are discussed.

After reviewing the literature on stimulus selection, Richardson (1971) noted the uniformity with which the data indicate a strong tendency for subjects to select the same type of component from all items. For example, if the stimulus compounds are made up of colors and words, the subjects tend to select all color or all word components (e.g., Underwood, Ham, and Ekstrand, 1962). If the compounds are made up of two CVCs or low-meaningfulness trigrams, the subjects tend to select components from a common position (e.g., Lovelace, 1968; Postman and Greenbloom, 1967). Many other examples could be cited, but Richardson's point is clear and well substantiated: the selected components can often be described by a single attribute or rule.

Subsequently, Dobbs and Carlson (1975) presented evidence that the selection of components is not only describable in this way but may actually be based on a common attribute or rule. If this is the case, what would happen if the learning materials were arranged so that each compound stimulus of a paired-associate list had a discriminative component but the attribute describing that component was not the same for every compound stimulus? The experiment to be reported used materials arranged in that fashion.

METHOD

Materials

The arrangement of the stimulus materials was patterned after that employed by Goggin and Martin (1970). Their stimuli were compounds composed of geometric shapes placed on distinctly colored backgrounds. Across trials, the shapes and colors were repeatedly re-paired such that the components from only one dimension were consistently paired with the same responses. Consequently, the components from one dimension were correlated with correct responding but the components from the other dimension were orthogonal to correct responding.

In the present experiment, the same arrangement was employed, with the notable exception that the components from one dimension were correlated with correct responding for only *half* the items and the components from the other dimension were correlated with correct responding for the remaining items. The list consisted of six items. For items 1–3, the three colors were consistently paired with the same responses, but the three shapes were continually re-paired from trial to trial. For items 4–6, three other shapes were consistently paired with the remaining responses, but the three colors were continually re-paired with these shapes. There were no attributes common to the three relevant colors and three relevant shapes (nor to the irrelevant three shapes and three colors) that could be used as a basis for selection. The responses (digits 1–6) were also paired in such a manner that they could not be used to group the relevant or irrelevant components. The components serving as relevant and irrelevant were counterbalanced with different subgroups.

Procedure and subjects

The subjects were read standard paired-associate instructions informing them of the type of responses and the bidimensionality of the stimuli, but not of the peculiarities of their arrangement. The materials were presented on a Stowe memory drum by the study/test procedure at a 2-sec rate. Four study and four test orders were used, each differing in how the shapes and colors were combined. Learning was continued for 40 study/test cycles.

The subjects were 8 male and 12 female students. They volunteered to participate as an option for course credit.

RESULTS AND DISCUSSION

Only 2 subjects learned the experimenter-defined list in the allotted 40 trials. The remaining 18 subjects rapidly acquired three items (mean trials to errorless performance for color or form relevant, 7.1) but 'failed' to acquire the remaining items.

Although a frequency count was not kept, subjects often reported that half of the list was easy but the remaining half took them a while to learn. These reports seemed somewhat at odds with the subjects' test records, as there seemed to be no evidence that all six items had been learned. Nevertheless, the subjects insisted they had. They explained that

three items always had the same response, whereas the remaining three responses were continually reordered on each trial. From their point of view, the task for the 'hard' items was to give the response that was paired with the experimenter-defined irrelevant component on the immediately preceding study trial. On the next trial, the pairing changed, and these new pairings had to be retained from the study to the test trial. When the data were scored in accordance with this complex task, 16 of the remaining 18 subjects did indeed learn the list.

Do subjects tend to select the same type of component from item to item? It would certainly seem that there is ample evidence they do. It seems astonishing that only two subjects selected components bidimensionally so as to learn the list in the experimenter-prescribed manner.

Despite the modesty of this experiment, the findings may have implications for statements of theoretical significance. As the data presented herein demonstrate, subjects have a strong tendency to encode all items of a stimulus set in a consistent manner. This is probably not an incidental observation. It means that items of a list are not learned in isolation and that theorizing at the level of the individual item may be inadequate. The suggestion is that the difficulty of discovering a basis for encoding in a consistent manner may in fact be a major determiner of the difficulty of encoding and, consequently, a major determiner of the rate of paired-associate acquisition. The presence of a prominent and differentiating attribute (e.g., color or form) should minimize the difficulty of encoding the stimuli and lead to rapid acquisition of the list when the attribute is correlated with the designated responses. If the prominent attribute is not differentiating or is uncorrelated with the designated responses, acquisition should be retarded.

The absence of a single prominent and differentiating attribute may retard the establishment of stable stimulus encodings. The search for a common basis for encoding may be responsible for the type of within-item encoding variability discussed by Martin (1968). However, from the point of view presented herein, it should be noted that Martin's assumption that the level of stimulus-term integration is a sufficient basis for predicting the amount of within-item variability is inadequate. The inadequacy stems from the focus of theorizing being at the level of single-item acquisition. When the entire *set* of paired associates is considered, it does not seem unreasonable to suggest that the difficulty of discovering and utilizing a common basis for encoding may be of more importance in determining encoding variability than the level of integration per se. In this regard, it may be more fruitful to focus attention on between-item variability — on the encoding of different attributes for different items. An

explication of those variables responsible for between-item encoding variability may prove more helpful in predicting stimulus recognition, acquisition, and retention.

Notes

This research was supported in part by Grant AO 335 from the National Research Council of Canada. The author thanks Beverly Davidson and Theresa Chrzanowski for their assistance in collecting the data and their helpful comments on the performance of the subjects. Requests for offprints should be sent to Allen R. Dobbs, Department of Psychology, University of Alberta, Edmonton, Alberta, Canada T6G 2E9. Received for publication February 18, 1976.

References

Dobbs, A. R., and Carlson, D. 1975. Transfer of stimulus-selection rules in stimulus coding. *American Journal of Psychology* 88:253–261.

Goggin, J., and Martin, E. 1970. Forced stimulus encoding and retroactive interference. *Journal of Experimental Psychology* 84:131–136.

Lovelace, E. A. 1968. Cue selection in paired-associate learning: Meaningfulness and position. *Psychonomic Science* 10:45–46.

Martin, E. 1968. Stimulus meaningfulness and paired-associate transfer: An encoding variability hypothesis. *Psychological Review* 75:421–441.

Postman, L., and Greenbloom, R. 1967. Conditions of cue selection in the acquisition of paired-associate lists. *Journal of Experimental Psychology* 73:91–100.

Richardson, J. 1971. Cue effectiveness and abstraction in paired-associate learning. *Psychological Bulletin* 75:73–91.

Underwood, B. J., Ham, M., and Ekstrand, B. 1962. Cue selection in paired-associate learning. *Journal of Experimental Psychology* 64:405–409.

AMERICAN JOURNAL OF PSYCHOLOGY
June 1977, Vol. 90, No. 2, pp. 313–318

Strategies in learning for recall and recognition tests

Joan Regan
University of California, Berkeley

The differential effects of studying for recall and recognition tests were studied by means of a transfer experiment. The 48 subjects learned three lists of paired associates with either all recall tests or all recognition tests and then were transferred to a fourth list with either the same test or the other test. Both recall and recognition on the fourth list suffered when the subjects changed tests. The data do not permit assigning the locus of the effect to encoding or retrieval.

Several studies have attempted to demonstrate differential learning strategies in anticipation of tests of recall or recognition, but the evidence so far has not been conclusive. With paired-associate learning, studies have found no effect, or an asymmetrical effect, of the two types of test (Freund, Brelsford, and Atkinson, 1969; Loftus, 1971). Carey and Lockhart (1973) found that recognition and free recall on a sixth categorized list differed as a function of the type of test used with five previous lists, but another study by Jacoby (1973) failed to confirm this finding.

Jacoby's subjects studied four blocked categorized lists with a single study and test trial for each. The type of training test on the first three lists (free recall, cued recall, or recognition) was combined factorially with the type of test on the fourth list. The data indicated no facilitative effect of training with free recall or recognition on fourth-list performance, as measured by number correct. Training with cued recall, however, did provide positive transfer, and these subjects performed better with cued recall on the fourth list than those whose training on the first three lists had been with either free recall or recognition. In this experiment, only the group trained with cued recall showed any evidence of learning to learn, as reflected in increased performance over the first three lists. Performance of the groups trained with free recall and recognition did not improve at all over lists. This suggests the possibility that one trial on each of three lists may not provide an optimal situation for the development of test-appropriate learning skills for free recall and recognition.

The present experiment elaborated on Jacoby's transfer design by providing more extended training with recall and recognition. The experiment did not try to pinpoint the locus of the effect of strategy in encoding or retrieval but only to demonstrate that strategy differs in learning for recall and recognition tests.

METHOD

Design

There were four groups representing the factorial combination of two types of *training* on three lists of paired words and two types of testing on a fourth list, the *transfer* list. The type of training involved learning three lists by a study/test method, each test being of recall (Rec) or of recognition (Rcg), the latter by a multiple-choice procedure. The same two kinds of test were employed for learning the fourth or transfer list, so that two groups learned it with the same kind of test they had had for the first three lists (Rec-Rec; Rcg-Rcg), and two groups changed to the other sort of test (Rec-Rcg; Rcg-Rec).

Materials

The materials were 92 one-syllable AA and A nouns selected from the Thorndike-Lorge (1944) list. Three lists of 10 pairs of words and two lists of 16 pairs were assembled. For any pair there was no rhyme or obvious association between the words, and each had a different initial letter. Presentation of the first three lists, each of 10 pairs, was counterbalanced so that within each condition each list was first, second, and third an equal number of times. The fourth list had 16 pairs rather than 10, to try to prevent a ceiling effect. Each 16-pair list was presented an equal number of times in each condition.

Procedure

Under all conditions, study and test trials alternated. The lists were randomly arranged into four different study orders and four different test orders. The rate of presentation was 2 sec per pair on the study trials and 5 sec per pair on the test trials. The interval between a test and study trial was approximately 3 sec. The first three lists were learned to a criterion of 9/10 correct, and the fourth list was presented for a minimum of five trials and continued after that, if necessary, until a criterion of 14/16 was reached. Under all conditions, subjects studied the pairs in silence.

On the *recall* tests, the subjects saw the stimulus word of each pair and tried to supply the correct response. They were encouraged to guess.

On the *recognition* tests, the subjects saw the stimulus word accompanied by four alternative responses. They were asked to read all four choices out loud and to repeat the word they believed to be the correct response. This was to ensure that they looked at all choices and to minimize the likelihood of their treating the situation as a recall test. The distractors were all response terms from within the list. This kind of test has resulted in subjects reaching criterion at a rate comparable to that of subjects learning with recall tests (Postman and Stark, 1969). The set of three incorrect alternatives accompanying a given

response changed from one test order to the next. Over the four test orders, each correct response appeared once in each of the four multiple-choice positions. For the 10-pair lists, within each test order each position was correct at least twice, but no more than three times. For the 16-pair lists, each position was correct four times. The interval between lists was 2.5 min.

Before the fourth or *transfer* list, the experimenter told all subjects that it would be the last list, what kind of test would be used, and that it would consist of 16 pairs rather than 10.

At the completion of the experiment and before debriefing, subjects completed a two-page questionnaire. For all subjects the *questionnaire* was exactly the same, and no mention was made of recall or recognition. They were asked to describe what strategies they had used, and if they had changed strategies, what the nature of the change had been and when it had occurred. The second page required subjects to rate the frequency with which they had used different encoding strategies for each list: association, imagery, visual and acoustic properties of words, plus an 'other' category. The questionnaire provided an explanation and an example of each labeled category.

Subjects

Forty-eight undergraduate psychology students at the University of California, Berkeley, participated in fulfillment of a course requirement. They were assigned to conditions by a randomized-block procedure.

RESULTS

Training

Performance on the first three lists was analyzed to include the type of test on the fourth list as a factor. This factor was not significant in any of the analyses. Table 1 displays the mean number of trials to criterion on the first three lists. There were no significant differences between the two groups trained with recall nor between the two groups trained with recognition, but learning was faster for the latter [$F(1, 44) = 10.46$, $MS_e = 2.98$, $p < .001$]. Although the expected equivalence in speed of learning did not materialize, both types of training improved performance over lists, with no interaction [$F(2, 88) = 23.13$, $MS_e = 1.53$, $p < .001$].

Transfer

The mean number of correct responses over five trials on the last list are presented in Figure 1. The first trial is the critical one for testing the differential-transfer hypothesis, and here the data clearly support it. Staying with the same type of test led to better performance than switching to a different type of test [$F(1, 44) = 9.09$, $MS_e = 9.99$, $p < .01$]. There was no interaction of training by transfer on this first trial, but over the five trials there was a triple interaction of training by transfer by trials

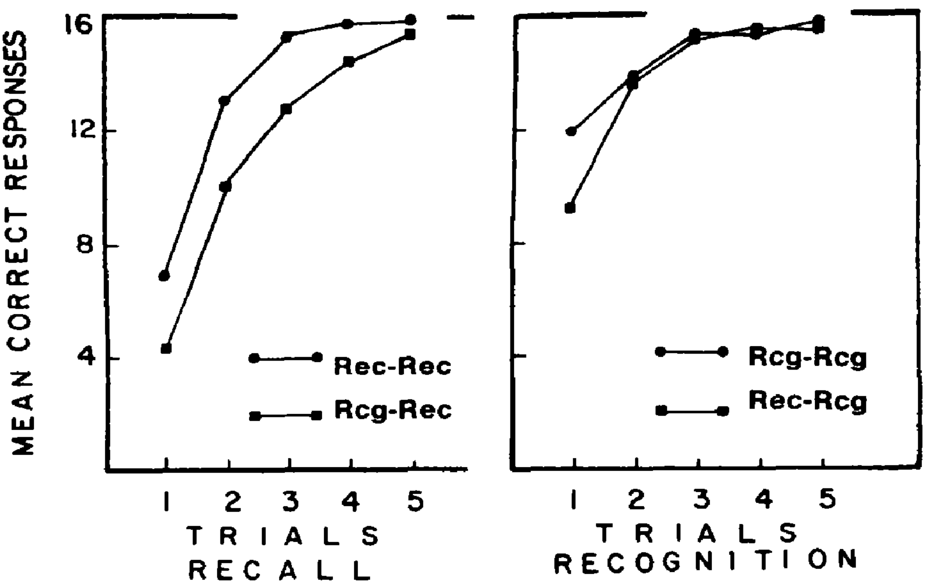


Figure 1. Mean correct responses over five trials on the transfer list

[$F(4, 176) = 5.11, MS_e = 2.30, p < .001$] and an interaction of training by transfer [$F(1, 44) = 8.26, MS_e = 12.13, p < .01$]. As the left half of Figure 1 indicates, the difference in the two groups tested on recall persisted through the fourth trial. On this trial, the F ratio was significant at the .025 level [$F(1, 22) = 6.75, MS_e = 1.39$]. Interpretation of the data on recognition, however, is hampered by an early ceiling effect on trial 2; see the right half of Figure 1.

The data on recall were rescored, but more leniently, so that any intralist response was counted as correct, even if given to the wrong stimulus term. This method of scoring did not eliminate the difference between the groups; it remained significant [$F(1, 22) = 4.71, MS_e = 9.69, p < .05$],

Table 1. Mean number of trials to criterion on the first three lists

Group	List 1	2	3	Mean
Rec-Rec	4.25	2 58	2.42	3.08
Rec-Rcg	4 75	3 33	2 92	3.67
Mean	4.50	2 96	2.67	3.38
Rcg-Rcg	3.25	2 08	2.08	2.47
Rcg-Rec	3 33	2 00	1.92	2.42
Mean	3 29	2 04	2.00	2.45

indicating that training with recall led either to an increased availability of intralist responses or to a greater tendency for this group to guess. The latter possibility seems unlikely, as both groups had a similar mean response rate (for group Rec-Rec, 8.25; for group Rcg-Rec, 8.75 on trial 1). As group Rec-Rec made so few errors, error percentages for the groups cannot be directly compared, but more than half of the errors by group Rcg-Rec were extralist intrusions.

Questionnaire

The data did not prove useful. All groups spontaneously reported using associations, imagery, and sentence formation in comparable numbers. Few reported changing strategy, and those who did claimed to have done so during the early lists rather than between the third and last lists.

DISCUSSION

In contrast to Jacoby's (1973) findings, the data indicate that subjects can take the requirements for recall and recognition tests into consideration as they handle the learning situation. Staying with the same type of test throughout four lists led to better performance on the fourth list than switching to a different type of test. The data, however, do not indicate the nature of the different skills involved. The patterns of errors in recall suggest that training with recall may lead to a better defined search set for those subjects (Rec-Rec). Training with recognition (Rcg-Rec) produced a considerable number of extralist intrusions on the first trial of the fourth list and even when these subjects were given credit for all intralist responses, they did not match the performance of the subjects trained with recall (Rec-Rec). Allen, Mahler, and Estes (1969), comparing the effect of study trials and test trials on long-term recall, also noted that experience with recall tests led to a decrease in extralist intrusions. However, LaPorte, Voss, and Bisanz (1974) have argued that the facilitative effect of recall tests cannot apply to nonspecific transfer. Their data indicate that when test trials are deleted during first-list learning, there is no detrimental effect on second-list learning.

While the present data do not determine the locus or relative importance of strategy in encoding or retrieval, they do clearly indicate that subjects can learn the appropriate strategy for recall or recognition tests. Training with recall in this experiment led to poorer recognition on the fourth list than training with recognition. This suggests that it may be misleading to contrast recognition as 'automatic' with recall as active and intentional (Kintsch, 1970).

Notes

The author thanks Dr. Geoffrey Keppel, who commented on an earlier draft of this paper. Requests for offprints should be sent to the author, Psychology Department, University of California, Berkeley, California 94720. Received for publication September 22, 1976.

References

Allen, G. A., Mahler, W. A., and Estes, W. K. 1969. Effects of recall tests on long-term retention of paired associates. *Journal of Verbal Learning and Verbal Behavior* 8:463–470.

Carey, S. T., and Lockhart, R. S. 1973. Encoding differences in recognition and recall. *Memory and Cognition* 1:297–300.

Freund, R. D., Brelsford, J. W., and Atkinson, R. C. 1969. Recognition vs. recall: Storage or retrieval differences? *Quarterly Journal of Experimental Psychology* 21:214–224.

Jacoby, L. L. 1973. Test appropriate strategies in retention of categorized lists. *Journal of Verbal Learning and Verbal Behavior* 12:675–682.

Kintsch, W. 1970. Models for free recall and recognition. In *Models of human memory*, ed. D. A. Norman. New York: Academic Press.

LaPorte, R., Voss, J. F., and Bisanz, G. 1974. Nonspecific transfer with and without test trials on the first list. *American Journal of Psychology* 87:475–479.

Loftus, G. R. 1971. Comparison of recognition and recall in a continuous memory task. *Journal of Experimental Psychology* 91:220–226.

Postman, L., and Stark, K. 1969. Role of response availability in transfer and interference. *Journal of Experimental Psychology* 79:168–177.

Thorndike, E. L., and Lorge, I. 1944. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University.

AMERICAN JOURNAL OF PSYCHOLOGY
June 1977, Vol. 90, No. 2, pp. 319–321

Notes from "Pavlov's Wednesdays": Sensory preconditioning

H. D. Kimmel
University of South Florida

Sensory preconditioning was first demonstrated in Pavlov's laboratory in 1931/32, rather than discovered by Brogden in 1939. Pavlov included nonassociative controls, forward pairing of the indifferent stimuli before reinforcing the second one with shock, and he avoided the development of inhibition to the compound by using a moving visual stimulus and a sound like that of scurrying mice, which both had persistent orienting reactions. Pavlov concluded that the indifferent stimuli were associated by temporal contiguity similar to human associations between successively spoken works. He did not consider the possibility of mediation via the orienting reactions.

American writers commonly credit Brogden (1939, 1947) with the first successful demonstration of the phenomenon known as *sensory preconditioning* (e.g., Kimble, 1961). Two indifferent stimuli, such as a light and a sound, are presented together several times, after which one of these stimuli is paired with classical reinforcement, such as food or electric shock. If the other stimulus subsequently elicits the response that was conditioned to the stimulus paired with reinforcement, sensory preconditioning is said to have been demonstrated. The phenomenon is considered important by learning theorists because it presents at least the appearance of a learned association based on nothing but the temporal contiguity of indifferent stimuli, without any apparent response or reinforcement being required.

The purpose of this note is twofold: to correct a long-standing historical error, since the phenomenon was studied and successfully obtained in Pavlov's laboratory in 1931/32, and to present Pavlov's reasons for doing the work and his recognition of the methodological complications it ininvolved. The topic is mentioned for the first time in Orbeli's volumes under the heading "Different Types of Conditioned Associations." It was October 21, 1931, and Pavlov wanted psychologists to see that there was a difference between their concept of association and his of conditioned reflexes. He suggested that N. A. Podkopaev do an experiment to "establish an association between two stimuli, which occasionally occur simul-

taneously, and then condition a food reflex to one of them; then [to] test the other stimulus, which has never been connected with food. It could thus be determined whether the other stimulus has also acquired conditioned food significance" (Orbeli, 1949, vol. 1, p. 156).

This procedure had been tried out earlier in Pavlov's laboratory, but with negative results. Pavlov suggested that Podkopaev use a variety of tones in conjunction with a light, in order to reduce the uniformity of stimulation and prevent the development of inhibition to the compound. After this first mention, the topic was not again addressed until Wednesday, December 21, 1932. The experiment Podkopaev had done did not involve food reinforcement or variable tones, from which it may be assumed that the procedure Pavlov had originally suggested was unsuccessful. Somewhere in the interim, there must have been a decision to shift to aversive conditioning, but nothing about this appears until after the data were collected.

In the seminar on December 21, 1932, Pavlov pointed out that human beings are able to form associations between even two completely meaningless words but that the conditioned reflexes formed by the dogs in his laboratory were "pragmatic," by which he apparently meant that the food or aversive reinforcement used in his laboratory were significant events to the animals and the conditioned responses biologically adaptive. Pavlov noted that a special technique is needed to determine whether a dog can associate two indifferent stimuli presented simultaneously. "Since the dog is unable to communicate with us regarding this, it was necessary to invent some kind of sign by means of which the possibility of forming this type of association could be established" (Orbeli, 1949, vol. 1, p. 262). Podkopaev simultaneously paired a whistle with a light several times. Then flexion of the limb was conditioned to the light, using electric shock as reinforcement. After this, the whistle also elicited flexion.

This work was followed up by I. O. Narbutovich, using two indifferent stimuli that elicited relatively slowly habituating orienting reflexes. These stimuli, a sound resembling that of scurrying mice and a rotating disc, were paired 21 times, with the sound coming on first and the rotating disc then being added. After this, the disc was paired with a weak electric shock to condition flexion to it. Then the sound was tested again and it elicited the flexion. As a control, a metronome was also used, but it evoked only an orienting reaction. Pavlov concluded his remarks on these studies by observing that, "thus, two indifferent stimuli were associated in this case exclusively by temporal contiguity" (Orbeli, 1949, vol. 1, p. 263).[1]

Pavlov returned to this topic only once more, on January 31, 1934. He was convinced that a genuine association had been formed between the

auditory and visual stimuli in Narbutovich's experiment, based on observations made even *before* the rotating disc was paired with shock. "When the sound came on, the dog not only showed an orienting reaction to the sound but also looked toward the moving light, and when the light came on, it reacted first in the direction of the sound and then turned to the light. It is clear that they had been associated with one another; this is the same association as occurs in us between two words that are spoken in succession." Pavlov's conclusions on the matter were that excitation was transmitted from the cortical center for the reception of the auditory stimulus to the center for the visual stimulus and, "since the light was associated with the defensive reflex, the sound also became associated with the defensive reaction" (Orbeli, 1949, vol. 2, p. 213).

Despite Pavlov's conclusions, clearly a stimulus/stimulus interpretation of association, his emphasis on using two stimuli whose orienting reactions do not habituate easily may provide some comfort to those who favor a stimulus/response interpretation whereby the association of the supposedly indifferent stimuli is mediated by the response they both elicit. The orienting reaction Pavlov wished to preserve obviously fits this interpretation.

Notes

Professor Kimmel currently is engaged in the first English translation of "Pavlov's Wednesdays," the series of seminars given by Pavlov in the 1930s. The present article describes an aspect of Pavlov's views that is important to both historical perspective and contemporary theory. We are pleased to publish this note, as we were its predecessors, in advance of the complete translation.—N.E.S. [Received for publication August, 1976.]

1. These experiments were subsequently described in detail in a research paper published in Russian by Narbutovich and Podkopaev (1936).

References

Brogden, W. J. 1939. Sensory preconditioning. *Journal of Experimental Psychology* 25:323–332.

Brogden, W. J. 1947. Sensory preconditioning of human subjects. *Journal of Experimental Psychology* 37:527–539.

Kimble, G. A. 1961. *Hilgard and Marquis' "Conditioning and learning."* New York: Appleton-Century-Crofts.

Narbutovich, I. O., and Podkopaev, N. A. 1936. Uslovnyy refleks kak assotziatziya [The conditioned reflex as an association]. *Trudy fiziologicheskoy laboratorii akademika I. P. Pavlova* [Proceedings of the academician I. P. Pavlov's laboratory] 6:5–24.

Orbeli, L. A., ed. 1949. *Pavlovskie sredy* [Pavlovian Wednesdays], vol. 1, *1929–1933*; vol. 2, *1934–1935*. Moscow and Leningrad: Akademiya Nauk USSR.

AMERICAN JOURNAL OF PSYCHOLOGY
June 1977, Vol. 90, No. 2, pp. 323–357

Book reviews

Carl P. Duncan
Northwestern University

Cognitive Theory

Edited by Frank Restle, Richard M. Shiffrin, N. John Castellan, Harold R. Lindman, and David B. Pisoni. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1975. Pp. 303. $16.50.

Between the periodic upheavals that Kuhn (*The Structure of Scientific Revolution,* 1963) has termed 'scientific revolutions,' there is a period of relative tranquillity called 'normal science.' Investigators work comfortably within the structure of their scientific paradigm. There is wide agreement on the appropriate themata and orienting concepts; there is little discord about the proper questions and appropriate evidence. The book under review is clearly *normal science,* in Kuhn's sense of the phrase. Its authors are excellent representatives of the information-processing paradigm in cognitive psychology. They neither decry other approaches (the word 'behaviorism' appears but once in the entire volume) nor self-consciously defend their pretheoretical ideas. The keynote is optimism: the emphasis is on harmony, and disagreement is minimized. Where disagreements exist, they are over details that empirical data collection is expected to resolve. Moreover, it is *early* normal science. That is, empirical research has not yet produced the stubborn anomalies that eventually will lead to the next revolution. Actually, the anomalies are present, if we knew where to look; but for the time being they are not bothering anybody. They exist as little ripples on the water, and only time will tell which ripples will dissipate in the experimental laboratory and which will swell into waves of dissatisfaction, bringing our next upheaval and our next paradigm. Finally, it is *good* normal science. Each author makes a small number of concise points, which are usually clearly stated at the beginning of his chapter. And all have emphasized the theoretical context of their work, rather than stringing together tedious lists of overinterpreted studies.

The book is the product of the Indiana Conference of 1974. It consists of four subsections — speech perception, judgment and decision making, short-term memory, and cognitive structures — each preceded by a brief overview by one of the editors.

In the area of speech perception, Studdert-Kennedy presents data suggesting that consonants are generally categorically perceived, vowels less so. He places these data in the context of a distinction between auditory processing (the processing of sheer sound) and phonetic processing (the processing of that subset of sound the hearer identifies as linguistic). Studdert-Kennedy argues that the

time span of vowels, which carry prosodic information, yields better auditory memory than consonants, whose defining features span so brief a time that they must be given rapid phonetic interpretation or be lost. The remaining three papers on speech perception offer techniques to study various aspects of the process. Cooper presents a method based on selective adaptation, along with the assumptions underlying it and samples of how it has been deployed. Briefly, it turns out that exposure to repeated syllables such as *ta-ta-ta-ta* will alter subsequent perception and production of other syllables in interesting and systematic ways. Wood presents data from a speeded syllable-identification task showing that subjects made redundancy gains when an auditory feature (pitch) was correlated with a phonetic feature (place of articulation). Wood interprets this to mean that auditory and phonetic features are processed in parallel, and he couches his explanation in terms of a decision-combination model. Pisoni demonstrates how variations of a dichotic syllable-identification task can explicate the feature-sharing advantage, by which dichotically presented syllable pairs are more accurately identified if they share such features as voicing and place of articulation.

One would be hard pressed to find any metatheoretical disagreement among these four authors. All regard the auditory/phonetic distinction as fundamental, and all are eager to get on with demonstrating the detectors and analyzers responsible for various aspects of speech perception. All find distinctive features, which are defined in articulatory terms, a suitable starting point; this suggests that all consider the relationship between perception and production to be an intimate one. All are interested in tracing the time course of mental activity following the input of sound. After reading the four chapters, one feels that we must know a great deal about speech perception — and yet, here and there one senses a ripple. As the techniques are pushed to yield answers to more and more specific questions, the answers seem shakier and shakier. The auditory/phonetic distinction seems well established, and the presence of specific systems for analysis of linguistic sounds seems adequately documented. But Cooper finds that the selective-adaptation effect is binaural; then further research yields a monaural component. For every postulated analyzer or detector an empirical counterpart is found; is it possible the techniques are creating them rather than discovering them? Some significant effects involve two or three milliseconds of time; not surprisingly, these effects are not altogether stable. Despite the wealth of detailed conclusions in these four chapters, many will require convergent validation by alternative experimental procedures before they attain complete persuasiveness.

The second section, on judgment and decision making, includes chapters by Birnbaum, Dawes, and Pitz. Birnbaum, who is concerned with psychophysical types of judgment, documents the roles of expectancy, context, and contrast in the judgments subjects make in the laboratory. To deal with this problem, he recommends the 'systextual' design, in which context is systematically varied. Dawes makes a strong case that much of our laboratory research has yielded a psychology of tasks rather than persons. The argument is that most psychological tasks have 'good' solutions, and the structure of these solutions constrains the adept subject. The paper by Pitz argues that Bayes' theorem is a useful model for the inference process. Pitz develops a way to measure at least some kinds of

belief, and he reports a finding that subjects take advantage of the opportunity to minimize memory load in a way that Bayes' theorem predicts.

The problems addressed by Birnbaum and Dawes transcend the areas of judgment and decision making; in fact, they pervade psychological experimentation. Subjects come to the laboratory and behave as they think they should in psychological experiments. Human processing of information is a remarkably general-purpose capacity. This means that subjects can rapidly develop strategies appropriate to the laboratory tasks, strategies which are not necessarily the same ones they use in the situations the tasks are intended to model. The problem of reconciling laboratory control with *sufficiency conditions* — the requirement that the theories be adequate to extralaboratory behavior — has no ideal solution. Birnbaum's systextual design may be appropriate when the relevant aspects of context can be identified and systematically varied, but this is seldom the case. Face validity in laboratory tasks is desirable but may sometimes preclude equally desirable analyticity. Perhaps the best solution is convergent validation of theoretical statements in a variety of laboratory situations. This approach is not infallible, but it does serve to protect against excessive scrutiny of our own analytical tools.

The section on short-term memory contains chapters by Bjork, by Craik and Jacoby, and by Shiffrin. Each presents a model of short-term memory, some of the relevant data, and a comparison to the other two models. Space precludes description of the three models here, yet taken together they give powerful hints where the field of short-term memory is headed. The authors address the similarities and differences among the three models, but what is more striking is the extent to which all three represent a departure from their predecessors. Shiffrin, who along with Atkinson (in *The Psychology of Learning*, vol. 2, ed. K. Spence and J. Spence, 1968) introduced one of the most influential of multistore models, now presents a new model in which both the sensory register and the long-term store seem to implode into short-term storage. The case against multistore models in general was made by Craik and Lockhart (*Journal of Verbal Learning and Verbal Behavior*, 1972, 11:671–684), so it is no surprise that Craik and Jacoby eschew the use of the term 'store,' although they do concede a primary memory that is different from the rest of the memory system. Only Bjork defends three separate stores; however, this may not actually be the most descriptive word for the components of his model. The authors are now more concerned with developing the interfaces of the stores than in documenting their differences; and for all three models, the component corresponding to short-term memory is not a place where items come and go but a state of the system at a particular point in time. While models of short-term memory may contain more than one component for a long time to come, the emphasis suggested by the term 'multistore model' appears all but dead.

A second notable characteristic of all three models of short-term memory is the extent to which their features are not dictated by the classic verbal-learning data. In fact, the main points of agreement among the three are in areas where intuition speaks more clearly than research findings. The models unanimously recognize that the things one is thinking about are different from those one is not thinking about; they all involve consciousness, attention, and working memory in an intimate relationship with one another, although the precise nature

of the relationship differs among the models. All contrast maintenance and elaborative rehearsal. And all incorporate Tulving's (1972) distinction between semantic and episodic memory; although the distinction is neither very central nor very tightly explicated in any of them, it apparently has great intuitive appeal despite its resistance to empirical validation. The three models all give central importance to permanent memory, or world knowledge — a concept that was aggressively ignored in the tradition of verbal learning and in early models of short-term memory. This probably does reflect the influence of laboratory data, though in a rather perverse way. By refusing to be interpretable otherwise, the data may have forced recognition of the intuitively obvious. In any case, the generality and interest of the three present models is greatly enhanced as a result.

The three papers in the last section are by Paris, Potts, and Restle respectively. Paris, using tasks similar to those of Bransford and his associates, demonstrates that even very young children go beyond the given, literal information to understand a prose passage. He also shows that this ability increases with age. Potts, using a task involving construction of linear orderings from prose, provides laboratory support for his claim that comparison times between pairs of linearly ordered items are an inverse function of the distance between the items. Restle interprets Heider's balance theory (*Journal of Psychology,* 1948, 21:107–112) in cognitive terms He also introduces a technique for studying how people organize themes to understand prose passages.

It is only in the last several years that the study of such evanescent capabilities as comprehension, reading for meaning, extracting themes, and so on has become an active area of research, even though these are among the most central problems for a psychology of human beings. Because no established methodologies have developed to stifle the imagination, each researcher is free to define his problem and mode of solution for himself. The three papers reflect very different approaches, all within the general information-processing mode: Paris addresses the developmental question, Potts researches a rather circumscribed problem analytically and in depth, and Restle uses rough-and-ready methods to grapple with broad, central questions.

In summary, this book is clear, well organized, and well edited. It is ideal for those already conversant with information-processing approaches to speech perception, judgment, memory, and comprehension. It defines no revolutionary new approaches but is a high-quality sample of a mature paradigm in operation.

Janet L. Mistler-Lachman, *University of Houston*

Psychological Research: The Inside Story

Edited by Michael H. Siegel and H. Philip Zeigler. New York: Harper & Row, 1976. Pp. 412. Paperback, $6.95.

There are nineteen chapters, all but two written especially for the book. The purpose of the book and the aim, more or less, of each author is to describe those aspects of research that do not get into the journals: the hunches, the blind alleys, the surprises, the amusing and depressing incidents, and occasionally, the serendipity. The authors' attempts to write informally succeed much of the time. Many chapters flow as easily as conversation; a few exhibit a massive formalism

of data and tables appropriate for the *Psychological Bulletin* perhaps, but not for a book such as this. Such chapters should have been more vigorously edited.

The book is divided into four sections. The first section, "Reflections on a Research Career," consists of two chapters, one by Harry Harlow on "Monkeys, Men, Mice, Motives, and Sex," and the other a paper by B. F. Skinner, "A Case History in Scientific Method," that originally appeared in the *American Psychologist* in 1956 (2:221–233). The second section, "Biological Foundations in Psychological Research," has four chapters. Howard Topoff writes on field and laboratory research on the social organization of army ants. Gordon Bermant deals with sexual behavior. "How to Turn a Rat into a Tiger: Environmental Determinants of Feeding Behavior" has four authors: George Collier, Robin Kanarek, Edward Hirsch, and Alan Marwine. The last chapter, "The K-Dogs" (after Kaspar Hauser, a boy presumably raised under extreme stimulus deprivation), is by John Fuller.

The third section, "Explorations in Experimental Psychology," has nine chapters. Two are on perception: one on "Seeing Is Deceiving," by Joan Gingus and Stanley Coren, and one on "Seeing and the Nick in Time," by Robert Sekuler. Three chapters are on cognitive topics: the meaning of spatial concepts (Lauren Harris and Ellen Strommen), memory (Lloyd Peterson), and hypothesis behavior (Marvin Levine). Stanley Schachter's previously published paper (*American Psychologist,* 1971, 26:129–144) is on obesity in humans and rats. Herschel Leibowitz summarizes experimental work on hypnosis, especially studies using visual perceptual tasks as the measuring tool. Dreaming and sleep research is covered by John Antrobus. Wayne Holtzman reviews the work on his inkblot test.

The fourth section, "Applications: Psychological Research and the Problem of Relevance," has four chapters. James McConnell writes of his own transformation from a researcher studying worms to a specialist in behavior modification and of his applications of techniques from the latter area to clinical psychology. Psychology in the courtroom and the reliability of witnesses is covered by Robert Buckout. Robert Sommer writes on the relatively new field of environmental psychology. Theodore Barber gives an introspective account of his own experiences under hypnosis and of his development of a training procedure for producing similar effects in others.

As can be seen, the book has enormous diversity. There is hardly a field of psychology that is not touched in some way. Diversity this great is usually deplored in reviews of multiauthored books, but I think it is one of the strengths of this book. Any student of psychology should surely find several of these chapters informative and even interesting. Different chapters would be worthwhile supplementary reading for different undergraduate courses. A hodgepodge of a book like this may not be to everyone's taste, but I found much of it fascinating.

C.P.D.

Walden Two, with a new introduction by the author

By B. F. Skinner. New York: Macmillan, 1976. Pp. 301. $7.95; paperback, $2.50.

"The early summer of 1945, when I wrote *Walden Two,* was not a bad time for Western civilization." So Skinner begins his new introduction. He then discusses the halting reception of his book at first and, after a while, its mounting sales,

increasing as compound interest would. "The world was beginning to face problems of an entirely new order of magnitude — the exhaustion of resources, the pollution of the environment, overpopulation, and the possibility of a nuclear holocaust, to mention only four." Is *Walden Two* the answer? Skinner's conclusion: "Something like a Walden Two would not be a bad start."

It will be assumed that the reader is familiar with the original book, a novel about a utopia describing a happy community whose design rests on Skinner's behavioral-engineering methods. In the new introduction, he deals with the woes of contemporary civilization and assesses the remedies he suggested years ago.

He uses primarily two arguments to justify the contemporary pertinence of *Walden Two*. The first is that behavioral engineering has come of age, and what was then science fiction has grown into "a technology of behavior . . . , no longer a figment of the imagination." The second is that the Walden Two he envisioned was small, and he believes we are now paying a high penalty for bigness. If we can reduce the size of the communities in which people live, we will be on the road to curing our social ills. "A network of small towns or Walden Twos would have its own problems, but the astonishing fact is that it could much more easily solve many of the crucial problems facing the world today." (All quotations are from the introduction.)

The human side of Skinner comes through very clearly in this new introduction. He wants all to have meaningful employment, to cultivate the best and most satisfying in the arts of leisure; he wants to reduce the use of alcohol and drugs, meet the issue of growing crime by eliminating its conditions and by rehabilitating criminals, establish a world without war or the threat of war, stop overconsuming and overpolluting. With these goals in mind, a utopian message is indeed appropriate, and it would be cynical to refuse to examine his remedies while accepting our woes as inevitable.

Skinner has confidence in his ideal community, but he does not tell us the experiences of the actually attempted Walden Twos and why they have not done any better than they have. As psychologists noting the decline of behaviorism and the emergence of cognitive psychologies, we do not all share his confidence in behavioral technology.

My own position is that we ought to have more pictures of utopia, not fewer, and should welcome Skinner's as one of them. Unless we have a vision of the ideal, our steps to improve the immediately possible may falter by creating as many problems as they solve.

Ernest R. Hilgard, *Stanford University*

Sensory Saltation: Metastability in the Perceptual World

By Frank A. Geldard. New York: Halsted Press, 1976. Pp. 133. $10.00.

This is a delightful little book. It offers two pictures: one of a fascinating sensory phenomenon, sensory saltation (the 'rabbit'), and the other of a famous investigator's happy efforts to pursue and tame that rabbit, as it comes to life in the book. The book was originally presented as three lectures, which served as the inauguration in 1975 of the John W. MacEachran Memorial Lecture Series at the University of Alberta. Professor Geldard's personal account of the discovery and investigation of his rabbit is charmingly revealing. As Geldard breathes life-

like qualities into the object of his pursuit, the reader discovers that it is Geldard rather than the rabbit who has been captured.

Sensory saltation is Geldard's formal term for the 'hopping about' of sensations that seems to occur between two or more spatially displaced stimulus sources when the sources are successively pulsed with certain temporal characteristics. The effect was originally discovered in Geldard's laboratory when an array of three vibrators on an observer's forearm was pulsed with a series of three successive trains of five brief square waves, one train to each vibrator. To the amazement of both observer and experimenter, the fifteen pulses seemed to dance up the arm in an evenly spaced sequence rather than to feel like three different trains at three separate points on the skin. That discovery, in 1971, has since been diligently pursued in an imaginative program of experiments that is described in this monograph. The experiments have been concerned with determining the necessary and sufficient conditions for the saltatory effect, with quantifying its dependence on spatial, temporal, and intensive variables, and with establishing its origin at a central rather than peripheral locus. Most of the experiments have involved the cutaneous modality, occasionally using electrical and thermal as well as tactual stimuli, but several in the auditory modality and many in the visual modality have also been conducted.

Although the precise interpretation and functional significance of the saltatory effect are at present uncertain, it seems usefully considered as an example of 'metastability.' In physics and chemistry, a metastable system is one in a state of pseudoequilibrium, with a delicately balanced organization that is readily modified by the addition of a small amount of energy from another source, thereby releasing a relatively large amount of its own potential energy. Geldard's thesis is that the illusory localization of stimulation in sensory saltation is an analogous case of metastable neural organization. Geldard's experimental data illustrate a systematic dependence of the spatial properties of sensory experience on an interaction of the temporal and intensive characteristics with the spatial characteristics of the physical stimulation. How the saltatory effect might be related to other sensory phenomena resulting from dynamic stimulation is not known, but Geldard points out that the subjective experience is qualitatively distinct from apparent motion or the phi phenomenon, arising as more compelling sensations at discrete points between the objective loci of stimulation. Geldard emphasizes that much remains to be learned about this intriguing phenomenon, and his monograph provides an excellent beginning to that study.

Geldard says that "this book is really the biography of a rabbit" (p. 12), but it is more than that. It is also part of a biography of a chaser of rabbits, a story of one who is captured by the object of his pursuit. Hopefully, many readers will also fall prey in a similarly blissful pursuit of Geldard's or of their own rabbits. Geldard's rabbit seems worthy game.

Joseph S. Lappin, *Vanderbilt University*

Principles of Learning and Memory

By Robert G. Crowder. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1976. Pp. 523. $19.95.

This new book, by Robert Crowder of Yale University, is the latest addition to the Experimental Psychology Series, a series with Arthur W. Melton as

consulting editor and published by Lawrence Erlbaum Associates. Like most of its predecessors, it is focused on human memory, learning, and information processing. Unlike most of its predecessors, it seems to be planned as a text rather than as a research monograph. As the author says in the preface, the book "grew out of lecture notes I developed over the years in both graduate and undergraduate versions of a course in human learning and memory at Yale." As a consequence, this book is a selective review of the field. While many topics are included, some are not, and those that are included are considered from a particular point of view. This is what one does in selecting and organizing material for a course, and given the proliferation in the field, what one must also do in writing any kind of review or survey. This is not a handbook; it is a selective review of the field of human learning and memory. Consequently, its evaluation will very much depend on whether you share the author's basic orienting attitudes and point of view. To the extent that you do, you will be favorably impressed with the book. To the extent that you do not, you will be disappointed.

The study of human memory has mushroomed in the past decade or two, and we have many theorists and experimentalists. Each side can list some impressive successes. At least in retrospect, we can say that Broadbent's filter theory in 1958 predicted the existence of a brief sensory store, documented only two years later by Sperling's findings on iconic memory. The paper by George Miller in 1956 on chunking by "the magic number seven" gave rise to what has loosely been called organization theory, with its plethora of experimental demonstrations and observations. Estes' notion of fluctuation in his stimulus-sampling theory has suggested the empirical phenomenon of reminiscence; while evanescent, there does seem to be some reliable evidence for the existence of this phenomenon. The buffer model of Atkinson and Shiffrin predicted a serial-position curve in long-term memory that would be monotonically decreasing from beginning to end, and the finding by Craik of what is called the negative recency effect in final free recall was impressive instantiation. Tulving has long argued for the importance of retrieval factors and has opposed the generate-and-edit view of recognition and recall. By his view, recall and recognition should be dissociable, and there might even be cases where recall would be better than recognition. Although still a contentious issue, there is some evidence for both of these propositions.

The experimentalists, by contrast, can point with pride to methods and procedures that yield reliable and reproducible results, sometimes (I can think of one) to the third decimal place. The classical example is Peterson and Peterson's discovery (published in 1959) that forgetting of a single item (really, three unrelated consonants) was essentially complete within a brief 20-sec period. One could trace out a complete forgetting curve for a single subject in a single 50-min experimental session. Proactive interference was impressively demonstrated as a phenomenon in long-term memory by Underwood in 1958, and again, a few years later, as a very significant factor in the short-term forgetting tested by the technique of Peterson and Peterson. Not only is there proactive inhibition, but release from proactive inhibition as well; extensive documentation has been provided by Wickens and his colleagues. Melton, like Underwood a major theorist of the interference school, has demonstrated that in repetition, the spacing between repetitions is important: quite unexpectedly

(and over a very wide range), the greater the distance (or spacing) between two presentations, the better the eventual memory.

This is a short list, on both sides, and it is not quite enough to fill a book. How and where to extend then becomes a question, and experimentalists and theorists would soon part company. The choice made in this book is quite clear. Crowder is firmly on the experimental side, not the theoretical side, and the contents of the book clearly favor data over theory. Obviously, theories and models are included: Estes' item-and-order serial model probably comes off best, but interference theory is carefully covered; the theories of Waugh and Norman, of Atkinson and Shiffrin, and the strength theories get a reasonable presentation; and even stimulus-sampling theory and signal-detection theory receive brief mention. But the bulk of the book is on data, and theories are secondary.

The author says in the preface that the book is organized into four broad sections. Chapters 2–5 deal with the question of coding in memory, and in particular, the relations between memory and vision, audition, and speech. Chapters 6–7 deal with short-term memory. Chapters 8–10 deal with learning, covering interference and transfer, effects of repetition, and organizational processes. Chapters 11–12 deal with retrieval processes, first recognition and then serial organization. The reader is advised to follow the ordering of topics within sections, but the ordering of the sections is somewhat less critical.

This book's three main themes are stage analysis, coding analysis, and task analysis. Stage analysis refers to the three phases of memory: the encoding or acquisition phase, which deals with the registration of information and, perhaps, the format of storage; the storage or retention phase, which deals with the persistence of or changes in the stored information over time; and the retrieval phase, which deals with the utilization of the stored information at the time of test. Coding analysis seems to be an attempt to deal with the representation problem; I don't understand how it differs from stage analysis of the encoding phase. Task analysis will be quite familiar to those steeped in the verbal-learning tradition. As an example, paired-associate learning can be broken down into stimulus recognition, response learning, and an associative stage, and other experimental tasks can likewise be broken down into specified components. These three themes (stage analysis, coding analysis, and task analysis) appear and reappear throughout the book.

In a general way, the book follows the organization common to many other books of its genre; namely, sensory memory first, here with separate chapters on iconic and echoic memory; short-term or primary memory second; and long-term or secondary memory last. This organization is not unreasonable if one believes in separate and distinct memory systems. Crowder considers this issue in the section entitled "Process Dualism in Memory." A popular view has been that short-term and long-term memory are separate and distinct, with different variables operating in the two cases. The contrary view denies that there are temporally distinct memory systems and argues for a continuum instead. Crowder clearly takes the first view, not the second, though he is careful to point out that his view is that a two-*factor* theory is necessary, not that there are necessarily two separate and distinct memory *stores*.

Like any book, this book has its strengths and weaknesses. On the positive side, it is clearly and carefully written, with great attention to detail. One is

lom in doubt as to what the author is saying. Many topics are hierarchically sented: first a general overview, then (not necessarily immediately) a de- ed exposition, and finally some concluding assessment. The experiments that discussed (and there are many) are generally discussed in sufficient detail t one has a very clear idea of their purpose, their method, their results, and ir significance. There are no 'one liners,' except as occasional signposts to ect the reader to special topics not covered in the book itself. Crowder is y careful to make it clear when he is giving his own opinion and when he being reportorial and to distinguish between what is orthodoxy and what is erodoxy. While there is much detail in the book, the factual density is not :rwhelming. It is not a book to read at a single sitting, but with spaced sentation one comes away with a healthy respect both for the book and for field. The book provides ample documentation that the field of human mory has come of age — that it has accumulated a solid core of experimental lings and has made a modest attempt at theoretical interpretation.

On the negative side, perhaps the book is *too* long on data and short on ory. Have we not reached the point (at least in some areas) at which spe- c experiments might yield general principles? Even better, might we not rt with a theory or model and discuss the experiments in that context, using theory, not the data, as the expository and mnemonic framework? The only ie this approach is really made is in the chapter on interference theory, and ny such accounts are already in the literature. In some other chapters, one s so involved in the details of so many specific experiments that it is easy ose sight of the forest for the trees.

have some reservations about this book, reservations not all would share. st, I was disappointed that the author does not distinguish between different *ds* of memories. It seems to me quite clear that different kinds of informa- ı are represented in memory. We have item information, or memory for cific objects and events. We have associative information, or simply associ- ons, and this includes not only associations mediated by temporal contiguity associations based on other relationships as well. We also have serial-order ormation, which underlies memory for serial strings that range from motor ls to the letters in the spelling of words. Item information, associative in- mation, and serial-order information are different. Their functional proper- vary markedly (far more than short-term and long-term memory seem to), l both the theories to explain them and the experimental paradigms to study m are different. I feel that any book that fails to emphasize such a basic dis- ction misses one of the main points to emerge from recent research in the a.

Second, I think it is too bad that the author accepted and featured the tem- al partitioning of memory. While we certainly have memories of different s (some are very fleeting, others endure), it seems highly unlikely that mory itself is organized or based on the duration of the trace. (Wickelgren's iew of the question is acknowledged, but not adequately discussed.) Not only he evidence not compelling, but the distinction between primary and sec- lary memory, even if correct, is logically unsatisfying. It should be deduced n more basic principles, not assumed *ab initio*. Ten years ago, it was reason- e to suggest this distinction, which has certainly been useful in organizing the rature and stimulating research. But what underlies this distinction, and why

does it exist? There are models in which primary and secondary memory are derived, not axiomatic, properties, but they are not included in this book.

Third, I think the organization is a bit weak. It is really organized on the basis of separate phenomena, which do not necessarily have a lot in common. Consider the chapter headings, which mention iconic memory, echoic memory, recoding by speech in short-term memory, nonverbal memory, primary memory, forgetting in short-term memory, the interference theory of forgetting in long-term memory, the effects of repetition on memory, the organization of memory in free recall, retrieval, and serial organization in learning and memory. With the clear exception of interference theory and the possible exception of primary memory, these are really all phenomena — effects to be explained. The 'organizing' themes of stage analysis, coding analysis, and task analysis are really not strong enough to tie these diverse topics together. Consequently, although the organization within chapters is exemplary, the organization across chapters is not all that it could be.

Finally, for whom can the book be recommended? Probably not for undergraduate students; it would be heavy going, too many facts and too few principles. It could be well recommended for graduate students; these students would certainly learn a great deal about the field, and although their instructor might find a few favored topics omitted, he would find much to applaud and little to carp about in the topics that are included. However, this book is probably best suited as a reference work for researchers in the field. If you are working on a topic Crowder has researched (and that includes many of the major topics in the field), this book is absolutely required reading. It gives a coherent, detailed, and thoughtful account of the experimental literature. It is here that the book makes its strongest contribution, and it will certainly be welcomed by many as a valuable and impressive commentary on the growing literature on human memory.

Bennet B. Murdock, Jr., *University of Toronto, Canada*

Enhancing Motivation: Change in the Classroom

By Richard de Charms. New York: Irvington Publishers, 1976. Distributed by Halsted Press. Pp. 279. $15.95.

The title of this book is both enticing and misleading. Therein, also, lies the source of frustration for this reviewer, for the book itself is characterized by the double thread of enticement and disappointment. To illustrate: the ambiguity of the title is evident after a few pages of reading, for it is not motivation in general that is to be enhanced. What the author and his collaborators (D. J. Shea, K. W. Jackson, F. Plimpton, S. Koenigs, and A. Blasi) describe is a program geared to enhance a specific process. Moreover, on several occasions the book is quite unclear on how this is a motivational process, and in one instance (p. 124) points out that the characteristic being enhanced is not specifically a motive.

The book has three main aspects. The first involves the author's concepts of Origin and Pawn, already introduced in an earlier book (*Personal Causation*) and in several articles. Another is the author's philosophy of science and his

concept of the nature of psychological processes and enquiry. A third, and the focus of the book, is a large-scale research project in an inner-city school district to train the children in the experimental groups to 'become Origins.' These three aspects are of quite different qualities. For example, the treatment of the concepts Origin and Pawn has some strong features, the concepts being useful, interesting, and referring to dimensions of human behavior that are not sufficiently attended to in contemporary psychological theory and research. Alas, the concepts as de Charms presents them lack precision both theoretically and methodologically. The second aspect of the book is probably its weakest: the author's philosophy of science can at best be viewed as muddled. One can sympathize with the author's hope to integrate subjectivism and objective behaviorism, but one must conclude that his efforts are not convincing and his rationale is limp. The overall lack of rigor in definition and clarification of concepts and methods is a major weakness of the author's whole approach. The third and focal aspect of the book, the research project that is described in a number of the chapters, appears to be a worthwhile first step in what promises to become a long-range systematic program for improving students' self-concepts and increasing the constructiveness of their efforts to deal with school and society. Some of the descriptions of the procedures followed in the project are delightful; many of the methods by which teachers, students, and researchers participated in the project are impressive for their human touch; and some of the procedures utilized in the project are innovative and intriguing.

The notion that some people are Origins and some are Pawns has a certain intuitive appeal. The author considers that individuals learn to be Origins or Pawns and that this is not a difference in learned ways of responding or perceiving but a difference in learned motivation. The defense of that position is tenuous at best. The characterization of Origins and Pawns alludes far more to a difference in attitudes, styles of responding, and concepts of self and of others. The author admits to difficulty in defining motivation, particularly because Origin and Pawn are not distinctions in terms of either goals or strength of motivation. Since the types of goals such individuals strive for are not clearly differentiated and since Origins and Pawns do not differ in explicitly described strength or intensity of motivation, it is an act of faith to accept the author's view that we are dealing with motivational distinctions. He is inconsistent in his use of the term motivation throughout the book, not only when referring to his own thinking but also to the writings of others, like McClelland. Since McClelland and his colleagues have offered quite explicit definitions and distinctions for terms like motive and motivation, this deficiency in de Charms' approach is not excusable. The author uses terms like goal, purpose, and commitment in loose and circular fashions, and he fails to deal systematically with the essentially cognitive domain inherent in his formulation of Origin and Pawn. A person who is an Origin is said to formulate his own goals; a Pawn, to follow goals set for him by others. An Origin is said to tend to pursue realistic goals and actions, to expect to be successful in attaining his goals, and to take responsibility for his own actions; a Pawn has opposite tendencies. These are basically cognitive processes, yet the author only superficially deals with Origin and Pawn as representing an individual's concepts of self and others. What reference there is to the literature on internal and external loci of control, attribution theory, or to Loevinger's formulation of stages of ego development

is fragmented and not tied to a well-articulated theory of the concept of self. Importantly, there is no single reference to Adler or Dreikurs, whose writings for the past 50 years have stressed that man is *active* and not *reactive* (terms de Charms uses, p. 4) and whose countless publications and video tapes have illustrated and described Adlerian counseling as a method of helping individuals feel that they determine their own actions and set their own goals! Although one would surely assume de Charms knows of the Adlerian literature and counseling methods, he makes no reference to them, even though he cites other and far less closely related work.

De Charms' approach to his research stems largely from the work of McClelland. McClelland's successful technique for training individuals to develop the achievement motive was the antecedent for the author's training program. In de Charms' program, the teachers were first trained to help them, and then to help them to help their sixth- and seventh-graders, become Origins. Special techniques were devised, many of them interesting and novel. Several chapters in the book give details of procedures that were used in the project to measure and evaluate the characteristics of an Origin. The reported results show that the training had some of the effects de Charms predicted it would on the children's orientations and academic achievement in some areas of school work. But it did not raise academic achievement in all areas nor even bring these students to mean achievement at national norms, as best as one can judge from the presented tables of results. The measure of the students' achievement motive was raised by the training, but not in all grades studied nor for both sexes equally. Important interactions were evident between de Charms' Origin training, achievement motive, and locus of control; a factor-analytic study would have been helpful. The correlations reported (especially in Table 7.5) for data obtained from the projective testing suggest that the fantasy measures of the effects of the training reflect a composite of achievement motive and locus of control and that the Origin training was in itself not a primary factor. The validity of the author's Origin/Pawn construct is still unclear, although the author makes an effort to come to grips with this problem in various sections of the book.

The reader of this book is likely to react to it in large part on the basis of professional background and interest. An educator will find the book to offer ideas that are interesting and stimulating and techniques that can be readily applied, to approach the classroom in ways that are fresh and creative yet direct and simple to follow, and to have the laudable intent to help children deal with school, peers, and society in a more confident and realistic manner. A psychologist concerned with motivation will find the book to leave large gaps. A psychologist concerned with personality and counseling will likely consider the book as attending insufficiently to the relationship between the Origin/Pawn construct, self-concept, and mental health; the Adlerian counselor especially will wish to see more emphasis on ways to build cooperation and stimulate social interest directly rather than as by-products of training the child to be more an Origin. A psychologist in learning and perception will probably not find the author's construct adequate, even though de Charms assumes learning and perception to be crucial to the development of 'being an Origin.' Thus, depending on the reader's background and concerns, this book will either provide an exciting new approach or a semifinished albeit interesting venture.

Eva D. Ferguson, *Southern Illinois University — Edwardsville*

Proprioceptive Control of Human Movement

By John Dickinson. Princeton, N.J.: Princeton Book Co., 1976. Pp. 209. $7.50.

Traditionally, the study of motor behavior has been the domain of psychology, physiology, and zoology, but in recent years, it has picked up kinesiology and sport psychology as partners. If these new partners are not actually in a college of physical education, they are a spin-off from it and are living nearby. Those of us in the sciences and humanities who have always associated a college of physical education with winning the engagement on Saturday afternoon and teaching our students the essentials of games would do well to update our perceptions. Dickinson's book on proprioception is part of this new thrust in scholarly research. The book was first published under the same title (Lepus, 1974) as part of the Human Movement Series, a set of British books on skill and sport psychology under the editorship of H. T. A. Whiting. The edition reviewed here is the same as the 1974 British edition.

Here are the chapters: "Historical Review," "Physiological Bases of Proprioception," "Measurement of Proprioceptive Sensitivity," "Proprioception and Performance," "Proprioception and Learning," "Proprioception and the Timing of Motor Responses," and "Proprioception and Training." The context of each of these is apparent except, perhaps, the one on timing. More often than not, theories of timing have been picturesque accounts of clocks in the head and, as might be expected, they have not generated much research. Exceptions are hypotheses about motor timing as a product of the characteristics of the proprioceptive accompaniments of movement, and these hypotheses have evoked enough research to deserve a chapter.

Among the book's positive features is its level of scholarship, both in the range of its coverage and in the description of individual experiments. Occasionally, the author includes criticisms and ideas of his own, which saves the book from being a listing of experimental findings — of abstracts chained together with weak transitional devices. The difficulty of the book is intermediate, which should be a helpful piece of information if one would like to consider it for a college course. Actually, the question is not likely to arise very often for undergraduate courses, because few of them engage the topic of proprioception. The book definitely has its place in graduate courses.

The negative features of the book are mercifully outweighed by the positive features, but nevertheless something must be said about them in the interests of a balanced review. A main shortcoming is the cursory treatment of the issue of the locus of motor control, a hot topic today for analysts of motor behavior. Is the regulation of movement fundamentally a matter of a central program that prescribes what the muscles should do? Or is the regulation basically determined by the peripheral feedback that is one of the consequences of movement, among which proprioceptive feedback is prominent? The debate has featured the manipulation of proprioception as an experimental approach, and Dickinson should have given greater coverage to it. Another deficiency is that the book generally restricts itself to human movement, as the title says. Physiologists and zoologists are astute students of the mechanisms of motor control and they of course use animals in their research, as do comparative psychologists who have done their share of relevant experiments. The human emphasis is understand-

able, given the roots in physical education that the book has, but some will find that emphasis slightly parochial.

Jack A. Adams, *University of Illinois at Urbana-Champaign*

Francis Galton: The Life and Work of a Victorian Genius

By D. W. Forrest. New York: Taplinger Publishing, 1974. Pp. 340. $14.95.

Many psychologists probably have the impression that Francis Galton did pioneering work in a variety of fields, including psychology. Let me admit that I did not know the half of it. Karl Pearson's biography of Galton is in three volumes and has 2,000 pages. It is easy to see where Pearson found that much material and hard to see how Forrest kept his biography to a more reasonable size. If most of Galton's 300 publications had been in psychology, rather than only a sizable fraction of them, his reputation as a genius, a master of many fields, might be less, but psychology would have benefited greatly.

Here are some of the things Galton did. His travels in unexplored regions of Africa, his 30 or so publications on travel and geography, and his many years as an active member and officer of the Royal Geographic Society greatly advanced both the exploration of the world and the acceptance of geography as a university department of research and teaching. He helped to put meteorology and weather forecasting on a scientific basis. He did work in experimental genetics, using rabbits and, later, sweet peas. (Yes, sweet peas. He did not, of course, know of Mendel's work.) He experimented with photography and developed the method of composite photography. He was one of the founders, possibly the major one, of fingerprinting as a method of identification. He invented a considerable variety of mechanical, optical, and other devices, including a rotary engine (100 years before the Wankel); an appendix in the book describes some of his inventions. And, of course, he published extensively on the inheritance of talent, on statistics, and on eugenics. The list could go on to his studies of association, of religion and self-induced religion, of self-induced paranoia, and so forth.

The book includes Galton's bibliography. From it one can see that his first publication (on his invention of a printing telegraph) was in 1850 when he was 28 years old. His last publication (dealing with the Eugenics Laboratory and the Eugenics Education Society) appeared in November, 1910, only two months before he died, at 89, in January, 1911. He was a scientist to the end. An hour before he died, when he could barely speak, he was given oxygen; he found voice enough to tell those around him that he had once experimented with the gas. An amazing and interesting man.

It is helpful to have available this biography, since most psychologists are not likely to read Pearson's massive work. However, after finishing Forrest's book, I was left with the feeling that I did not really get to know Galton. Forrest does mention on several occasions that Galton was a hard man to know; in some ways even his wife did not seem to penetrate his reserve. To a great degree, Galton's life was inside his very large head: a good thing for science but tough for biographers.

C.P.D.

Psychophysics: Method and Theory

By George A. Gescheider. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1976. Pp. 177.

Students of psychophysics often cite 1860, the date Fechner's *Elemente der Psychopysik* was published, as the birth of their discipline. But historians of the future may consider 1953 an even more significant date, it being the year of announcement of not one but two ingenious ideas that have together transformed and revitalized the study of psychophysics. In that year, S. S. Stevens published an abstract in *Science* (118:576) that marked the beginning of his program of research on the scaling of subjective magnitudes. The result was a sharp break away from the traditional preoccupation with threshold phenomena and toward a measurement of the response to suprathreshold intensities — toward the displacement of Fechner's long-lived logarithmic law by the psychophysical power law. In the same year, Wilson P. Tanner, Jr., and John Swets announced a "new theory of visual detection" in a technical report from the University of Michigan, one of the first applications of the theory of signal detection to a problem in psychophysics.

Since 1953, the separation and independent measurement of sensitivity and bias in psychophysical judgment has been firmly established, and a permanent change in both the theory and measurement of thresholds has resulted. Until now, the teacher who wished to introduce these developments to undergraduate students without being superficial was obliged to assign primary sources that, even with extensive classroom explication, were still rather heavy going. Now Gescheider, in just 163 pages of text, has provided us an excellent summary of the main ideas and data associated with the psychophysical power law and the application of signal-detection theory to psychophysics.

To provide a background, Gescheider begins with two chapters devoted to classical Fechnerian psychophysics: absolute and difference thresholds, Weber's law, Fechner's law, and the classical methods of psychophysics. It is a standard and straightforward treatment, well within the capability of undergraduate psychology majors, but I fear that it may reinforce the mistaken notion that this part of psychophysics is simply history and nothing more. For example, Weber's law is stated, along with the classical correction of it, but no mention is made of the modern data on the 'near miss' to Weber's law. Again, I think it would have been wiser, and more helpful to the student, to give more space and detail to the staircase and forced-choice techniques of threshold measurement, which are currently in wide use, than to the method of constant stimuli, which now rarely appears in the pages of our journals.

A third chapter is concerned primarily with an exposition of the theory of signal detection as applied to psychophysics and as compared to classical threshold theory and the theory of the neural quantum. *ROC* curves, *d'*, and basic methods are all carefully and cogently described. Perhaps the greatest emphasis, however, is placed on the ***implications*** of signal-detection theory for the concept of threshold; that, I think, is a mistake. First, it may leave the student with the impression that these implications are the principal significance of signal-detection theory in psychophysics. But once the high-threshold model had been eliminated, there emerged a variety of more sophisticated and complex threshold models capable of generating smooth *ROC* functions indistinguishable from those

predicted by signal-detection theory. Indeed, since Krantz's 1969 paper showing the equivocal nature of evidence that had been regarded as intolerable to threshold theory, the whole issue has declined substantially in popularity. Second, as a result of this emphasis on the implications of signal-detection theory, little room is left to describe the many ways in which the disentanglement and separate quantification of sensitivity and bias has produced new knowledge. The discovery that the decline in performance during extended vigilance tasks is almost entirely due to a change in criterion and that sensitivity remains invariant is but one example. More emphasis on the *empirical accomplishments* of signal-detection theory in psychophysics would have been most welcome, especially in a text for undergraduates. However, Gescheider does earn my thanks for restating and developing Krantz's distinction between observer thresholds and energy thresholds. The former involve a question about the internal state of the observer, a question that has been a target of study by signal-detection theorists, whereas the latter may exist quite independently of the answer to that question. Gescheider reports some of his own data demonstrating a range of stimulus energies (vibrotactile stimulation of the finger) for which d' is zero and above which d' increases with energy. Here is unequivocal evidence of an absolute threshold that is quite independent of the argument over the continuity or discontinuity of internal states.

The last two chapters have to do with psychophysical scaling and Stevens' psychophysical power law. Once again, there is a clear and careful development of the subject; it should pose no difficulties for a capable undergraduate. My only serious objection (and I know many will not share it) is to the thoroughgoing dualism embodied in the author's use of the language of sensation. Beginning with the opening statement in the preface that "psychophysics is the scientific study of the relation between stimulus and sensation" and continuing to later statements such as the one that "in using direct scaling . . . , the observer makes judgments of his sensations which are then directly converted into measurements of sensory magnitude" (p. 86), Gescheider presents psychophysics as the Fechnerian enterprise of relating mind and matter. Gescheider does repeat Stevens' rebuttal to this criticism as being more a matter of words than data but leaves me (and, I suspect, would leave many students) with the impression that the method of magnitude estimation provides a direct measurement of sensation and that the resulting power law concerns not only an empirical relation between stimulus intensity and numerical judgment but between stimulus and sensation. The latter is in my view a matter of theoretical speculation and should be identified as such to the student reader. And, given my bias, I would have preferred to see Stevens' scaling methods defined in terms of cross-modal matching with magnitude estimation developed as only a special case.

Finally, I could have wished for a textbook that posed more questions for the student to worry about. For one example, the difference in results obtained by magnitude estimation and by magnitude production continues to puzzle investigators and to provoke new and interesting research; that might not be guessed by a student reading Gescheider's treatment. Or again, his account of category judging as a scaling technique could have been written 15 years ago and gives no inkling of the sizeable body of research on category scales that has accumulated since then. For example, the extensive use of this method by Norman Anderson is not even acknowledged to exist. Further, I think it is a mistake

to dismiss category scaling, as Gescheider does, for failure to meet the requirement that "the judgment should be completely independent of the values of other stimuli presented on other trials" (p. 108). When this objection was first raised (by Stevens and Galanter in 1957), it was thought, mistakenly as it turns out, that magnitude estimation *did* have the desired freedom from contextual constraints. Even though we now know that it does not, few have suggested that we therefore abandon it. Students should be aware that category scaling continues to be widely practiced and that the relation of the results to those obtained by magnitude estimation continues to be the subject of debate.

But most of these flaws are easily remedied by an alert instructor and are minor in significance compared to the general success with which the main purpose of this volume has been met. That purpose: to summarize the contributions of the theory of signal detection and of Stevens' scaling procedures to the field of psychophysics and to do so at a level within the grasp of an undergraduate. Here, then, is a good textbook introduction to two of the most fruitful ideas to emerge in the field of experimental psychology. Every psychology major should read it.

Robert Teghtsoonian, *Smith College*

Human Learning

By David L. Horton and Thomas W. Turnage. Englewood Cliffs, N.J.: Prentice-Hall, 1976. Pp. 501. $12.95.

Textbooks in the psychology of learning have traditionally presented this most important of human psychological activities in terms of the data obtained from the study of nonhumans. The discussion of human learning has been conducted within the context of theory derived from experiments with rats, pigeons, and sometimes monkeys, with specific reference to research on human learning only when dealing with a few topics traditionally studied in work on verbal learning.

Traditions are changing, however, and Horton and Turnage's text reflects this change. As a result, the text differs considerably from those currently available. The approach to learning based on conditioning principles is given short shrift, and the approaches in terms of information processing and cognitive theory are stressed. The book is atypical in another way. Rather than attempting to integrate the available literature, it emphasizes major theoretical trends and issues, with the argument proceeding by specific theoretical and empirical examples instead of exhaustive citation.

In the late 1960s, there was a notable increase in the number of psychologists who seriously questioned the generality of theories based on simple laboratory paradigms and nonhumans and who began to consider special and perhaps distinctively human processes in learning, particularly language. At another level, this text describes the systematic and intellectual forces leading to this change both at the theoretical and empirical level.

The actual content is wide-ranging. The text is divided into three sections. The first section is on what Horton and Turnage call the learning approach. It deals with the two traditional research areas concerned with learning: conditioning and verbal learning. The second section is theoretically oriented toward information processing and particularly human memory: the storage

and retrieval of information. The third section is somewhat of a hodgepodge characterized by the authors as the structural-analysis approach. It includes a variety of linguistic phenomena and concept learning. The common thread is that contemporary approaches to the explanation of these phenomena typically make reference to complex relations among elements, the use of which is governed by rules.

With few exceptions each chapter is virtually a new topic and while there is a sort of logical and historical linearity to the presentation, it is perhaps most instructive to consider the chapters one at a time. The introductory chapter wrestles briefly with the problem of defining learning before tracing the historical antecedents of a number of pretheoretical biases (e.g., nativism/empiricism) and key concepts through the last 100 years. Pavlov and Thorndike are considered in some detail, but the focus of the chapter is a critical analysis of the emergence of behaviorism and the theory of learning it spawned.

The first section begins with two chapters on conditioning that are really quite different from each other. The first chapter is actually a continuation of the analysis of learning theory begun in the introduction. The basic laws of association — frequency, contiguity, and effect — are discussed within the framework of Thorndike's early research. A brief discussion of conditioning paradigms and some of the major phenomena is followed by two examples of the application of conditioning principles to more complex phenomena: Spence's 1936 theory of discrimination learning and the concept of mediating responses. The discussion is openly critical of the approach to learning theory through conditioning. If one were to read only this text, it would be difficult to see whether this criticism is deserved. The need for mediating events, the necessity of specifying the stimulus, and some of the other problems raised by the authors do not in themselves provide a compelling reason for rejecting this approach, and generally a sophisticated 'behaviorist' can easily counter them. The problem is that no one can realistically deal with the philosophical complexity of this issue in so few pages. In attempting to do so, the authors have oversimplified many issues and failed to give readers a fair picture of the potential of a general behavior theory.

Some of the omissions are serious. For example, no distinction is made between behaviorism as a methodological tenet and as an epistemological position. Nor is it made clear that psychologists treat the conditioning experiment in quite different ways. While some do consider it a direct source of universal empirical principles, many others quite clearly consider it merely a crucible for formulating and testing theories. While such theories have traditionally been based on associations of stimulus and response, many newer theories talk about *contingencies,* which sounds suspiciously like information processing. For a student who has not previously studied conditioning, the basic phenomena are not adequately described. The presentation is less comprehensive than that found in many introductory texts.

In contrast, the second chapter is largely empirical. Some studies of the classical and instrumental conditioning of humans of all ages are described. Included is a discussion of the issue of awareness in verbal conditioning. The remainder of the chapter covers the application of conditioning principles to a wide range of real-life situations: pathology, social behavior, animal training,

and education. While the empirical effectiveness of the procedures is difficult to deny, the authors often question the theoretical relevance of the procedures both as corrective devices and as explanations of how various pathological phenomena arise. The arguments are the standard ones and are well put, but again the issues are complex and the opposite side of the debate is not fairly stated.

The final two chapters in the first section cover verbal learning: one on associative learning and the second on forgetting. After an excellent summary of Ebbinghaus' research, the chapter on associative learning proceeds to the discussion of a number of traditional problems in the learning of serial and paired-associate lists. The serial-position function, remote associations, and the nature of the stimulus are the problems in serial learning that are considered, while the effects of meaningfulness, multiprocess theory, backward associations, and the all-or-none issue are dealt with under the paired-associate rubric. The chapter on forgetting is devoted almost entirely to interference theory based on studies of retroactive and proactive inhibition. There is altogether too brief a discussion of transfer, in which the theoretical relation between transfer and retroaction is not considered and little data is presented.

It is difficult to comment on these chapters alone because their subject matter is so clearly interwoven with that to be presented three chapters hence. The study of verbal learning and memory has expanded both theoretically and empirically and has done so in an orderly rather than revolutionary manner, with theoretical change closely following the demonstration of phenomena that seem to demand the change. Because the theory is in a state of continual flux, it often has no 'traditional' form. Systematically, verbal learning has been functional in its commitment to the concept of association rather than using it as an explanatory process. Given this state of affairs, there are a number of puzzling aspects of the presentation. Certainly, one could not find better examples of conceptual and empirical confusion than the stimulus problem in serial learning, the backward-association problem, or the effects of *m*. If the authors are trying to show that the failure to solve these problems augurs theoretical change, they have succeeded.

On the other hand, notable omissions lead to an erroneous impression of the conceptual range of the field. For example, research on stimulus selection is certainly as important to verbal learning as is work on backward association. Martin's theory of encoding variability is not even mentioned here, despite its implications for acquisition, transfer, and interference theory. Nor is the frequency theory of verbal discrimination considered. None of these deals exclusively with associative processes, and yet they have played a prominent role in the research on verbal learning in the 1960s. The inconsistency between the all-or-none problem and a multiprocess analysis is not even noted.

The second major section of the book consists of two chapters on memory sandwiched around one on attention. The first chapter is based on the dual-process theory by which memory is conceptualized as two separate systems: short-term and long-term memory. The data presented are designed to bear on this notion and are mostly from experiments on short-term or immediate memory. There is a discussion of coding processes with reference to Wickens' research, the semantic/acoustic problem, and Underwood's attribute theory.

Finally, three specific models are presented at various places: those of Waugh and Norman, of Atkinson and Shiffrin, and of Craik and Lockhart.

The authors seem convinced of the veracity of some form of dual-process (or multiprocess) theory. They cite its critics (without bearing down on the criticism) and do not offer Craik and Lockhart's most cogent arguments for an alternate approach. The selectivity of the data results in a number of important phenomena being ignored. The most salient of these are the rehearsal effects, which are important to Craik and Lockhart's view and to an explanation of the lag effect. Horton and Turnage present the encoding-variability interpretation of the lag effect as the only one, when rehearsal and other processes are much more likely to be important.

The second chapter in this section is on attention and only tangentially deals with memory. There is a brief historical introduction and an even briefer look at the physiology of stimulus selection before the major thrust of the chapter appears. The well-known studies of visual iconic memory are presented, followed by a consideration of a possible auditory counterpart. There is a short discussion of the development of attention and limited capacity, followed by a description of Sternberg's memory-scanning paradigm. The chapter closes with an extensive look at divided-attention research and filter theory.

The third chapter returns to memory. It discusses three major topics: organization in free recall, coding processes, and semantic memory. The discussion of organization deals with a number of phenomena (e.g., clustering, frequency effects, context effects) but is highly selective and ignores the theoretical issue that has captured the most attention: generation/recognition theory and the cuing problem. The theoretical part of the discussion of coding draws on Posner and Warren, Bower, and Paivio. In addition, several classic and illustrative studies are cited. The discussion of semantic memory sets the problem and then describes three models: those of Quillian, of Rumelhart, Lindsay, and Norman, and of Anderson and Bower. Associative approaches to the problem of semantic memory (e.g., that of Deese) are scarcely mentioned.

The final section of the book is represented by four chapters. Three of these deal directly with speech. The first chapter is concerned with the physical signal, notably, an analysis of the relation of phonemes to various psychological functions. The topics dealt with are speech perception, brain lateralization, the nature of the speech unit, rhythm, and acquisition. Each topic is based on one or two authoritative reviews.

The second chapter is basically linguistic, and at that, restricted to a consideration of syntax. The authors implicitly accept the notion of a generative grammar and deal effectively with the distinction between performance and competence. There is a listing of the typical competencies cited by some linguists as being psychologically relevant and a description of finite-state, phrase-structure, and transformational grammars. The last part of the chapter describes some of the research relating various grammatical structures to behavior.

The third chapter is a potpourri of diverse topics ranging from language acquisition to comprehension. The section on development is highly speculative and theoretical. The remainder of the chapter is a concatenation of empirical findings with specific experimental paradigms. The focus is on the sentence, and

the problem is how sentences are 'understood.' The theoretical ideas range from two versions of deep-structure theory based on syntax (derivational complexity and verb complexity) to a consideration of Clark's 1969 model based on semantics. The empirical paradigms are sentence memory, phoneme monitoring, and Bransford and Franks' inference paradigm.

It is difficult to see why most of this material was included in a book on learning. None of the three chapters deal even indirectly with the learning process in any detail, even granting the authors' initial bias to consider primarily verbal behavior. They tell us little either about how we learn to talk or understand verbal speech or about how this ability is involved in the learning of other things. The chapters, moreover, are badly dated, both in terms of linguistics and in terms of the issues in psychological research presently extant. Historically, they manage to capture the flavor of psychology's first serious encounter with formal linguistics, but as usual, psychologists have gone their own empirical and theoretical way, and formal linguistics is no longer of much relevance to psychological theory.

The final substantive chapter is on concept learning. The empirical point of departure is the attribute-classification task. The chapter largely consists of a parade of specific theories, beginning with stimulus/response and mediational theories, proceeding to the hypothesis-testing theories of Bruner, Goodnow, and Austin, of Restle, of Bower and Trabasso, and of Levine. Several issues faced by these theories are also considered in reasonable detail. Considerations of information-processing approaches as exemplified by Hunt's model, Bourne's truth-table approach, and an inductive model authored by Trabasso, Rollins, and Shaughnessy precede an extended discussion of a little-known theory of Jenkins.

The last chapter is a review. It concentrates on the demise of stimulus/response theory but also comments on other theoretical approaches.

While I have attempted to make certain critical points about individual chapters, what can be said about this book as a whole? There will probably be many different reactions to this book depending on the readers' expectations and theoretical biases. Some instructors will no doubt be pleased by a text that challenges the relevance of conditioning to the study of human learning. Others will be amazed to find their entire area of research dismissed as irrelevant on what they perceive to be immaterial grounds. Many will certainly disagree with Horton and Turnage's interpretation of history.

My own reaction was not generally positive. As a basic text for students beginning the study of human learning, it is too selective and polemic for my tastes, plus the fact that about a quarter of the book does not deal with learning at all. Moreover, I am not convinced that the selection is fair or that the systematic arguments are relevant to the mainstream of psychological research. Metatheoretic analysis is a complicated business, and it is at this level that 'approaches' must be evaluated. While the authors make statements of a metatheoretic kind, they do not pursue them at a depth that would make their arguments logically or empirically persuasive. Their conclusions may be correct, but this book does not adequately document the reasons for those conclusions.

Students who have had a preliminary course in learning from the point of view of behavior theory may benefit from this text, which clearly presents an

opposite point of view. Without that background, the most important issues are likely to escape them.

Willard Runquist, *University of Alberta, Canada*

Logical Abilities in Children
Vol. 3, Reasoning in Adolescence: Deductive Inference

By Daniel N. Osherson. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1975. Pp. 272. $15.00.

A minimal expectation of a book claiming to study logical abilities is that it be logical and well reasoned. Osherson's third volume on logical abilities in children meets this expectation; indeed, exceeds it, in that a good deal of the book deals with the logical process from each minute step in the theory-building process to the next. This highly detailed treatment of logic in the adolescent consequently makes significant demands on even the serious student of the topic, and extraordinary demands on the more casual observer. Several comments on the style of presentation will follow a consideration of the substance of the theory and the data it generates.

Osherson's overall question for this series is the central one in cognitive development. How can one decide whether the thought or, more specifically, the mental structures of the child, the adolescent, and the adult are similar or dissimilar? In a lucid introduction, the author reviews all the problems inherent in obtaining an answer to this central question. He examines the Piagetian position, which argues for major dissimilarities; justly criticizes it for failing to provide reasonable empirical verification of the hypothesized dissimilarities; and concludes that "the issue of qualitative discontinuities between children and adolescents in logical reasoning is still open, despite the obviously important work of Piaget" (p. 8).

What then is the basis for decision? Employing a useful analogy to different dialects of the same language, the author argues that "we should judge the underlying thought processes governing these (logical) abilities to be similar if and only if the theories accounting for the abilities at the two ages are similar" (p. 9). The core of the book is an attempt to construct a theory of the logic of adolescents that will provide the basis for comparison to a theory of the logic of children or of adults.

In fairness to the prospective reader, it should be noted that the author offers something less than strong encouragement to those who might travel with him that the voyage will be ultimately worthwhile. In the preface, Osherson concludes with the candid statement that the "model developed in this book is innocent of concerns that I now consider central to the study of deductive reasoning. Discussion of these matters begins in Sections 21.2 and 21.3 and resumes in Volume 4." This follows a pattern in the series: the outcome of volume 1 was eschewed in volume 2. While perhaps setting a record for flexibility in psychological theorizing in response to additional evidence and problem finding, this may also lessen one's desire to part with $15.00 for volume 3 when the central issues are to be found in volume 4. A wiser course might have been to follow more traditional routes of publication (i.e., journals and monographs) for such theoretical development and empirical testing.

It would be unfortunate if this work were to receive less exposure for so ·eripheral a reason, because this is a highly educational exercise in theory ·uilding and testing that might possibly lead to clearer thinking in the currently onfused area of adolescent cognitive development. The author narrows his onsideration of adolescent logic first to the problem of deductive inference and hen to the processing of intuitively valid arguments rather than proof finding. ;ach of these restrictions is arbitrary, but each is made for good methodological easons and serves to focus the discussion. In addition, the model is built to ,ork for propositional logic, but with the hope that it will be generalizable to ther logical statements, especially class inclusion. Finally, a restriction that the nodel be one of actual cognitive processing is also imposed.

Of three possible approaches considered in this book, logistic (i.e., formal ogic), semantic (i.e., truth functionality), and natural deductive, Osherson hooses the last as the most appropriate beginning point. The theory consists of a et of operations that will make accurate predictions, it is hoped, of subjects' ctual performance in the evaluation of the validity of arguments. The basic ·aradigm for the collection of evidence is to provide adolescents (who are some- ,hat casually selected, the model being intended to be universal and thus un- oncerned with individual differences) with a set of arguments that differ ı specifiable ways and ask them first to decide the validity of the arguments nd second to rate the difficulty of the arguments they judged to be valid.

Two major criteria are used in each experiment to evaluate the performance f the model being examined. First is the *inventory* requirement, which stipu- ıtes that a subject who rejects an argument composed of several of the single ·perations proposed by the model must also reject at least one of the single perations included in the multiple-operation argument and must not reject ny of the single operations included in an accepted multiple-operation argu- ıent. Second is the *additivity* requirement, which states that the difficulty of a ıultiple-operation argument (measured here by rating scales) should be some near function of the difficulties of the single operations that comprise it. Vithin-subject consistency is a minor consideration for the author, although etween-subject consistency across two sessions is important.

Osherson begins with a model introduced in volume 2 and proceeds to revise nd refine it, based somewhat on the evidence from the studies reported here, ;pecially that on the inventory and additivity requirements, and based also on desire to improve the range of the model. The results are generally promising ›r the group data but are discouraging at the level of individual prediction, 'hich is of more importance than the author allows when a model of real-world ıental processing is being argued. On the group data, for example, the correla- ons between average rated difficulty of multiple-operation arguments and the ımmed average difficulties of its component single-operation arguments hover round +.90. For individuals, however, the correlations are much lower, around -.30 to +.40. The author attributes this to 'noise,' and undoubtedly much of is. But a true model of logical processing would seem to require stronger erification and less rationalization. The outcome is that the model, while romising, is still a highly tentative one.

Two general difficulties should be noted as well. The first is that the model ıay be unnecessary for a psychological theory. Since the derivations the model enerates are finite, they could be listed as the theory. A generative model may

be more simple, but unfortunately a number of models can generate the same derivations. Further, since it is the derivations that provide the empirical test, there is in this work no direct test of the prior model. The second difficulty is that Osherson now sees the semantic component as a logically prior issue for a natural-deductive model.

For those who have the time and interest to devote their efforts to a careful study of this book, the payoff in clarification of issues is considerable. The author's style is both highly succinct and technical, requiring study rather than casual reading. On some occasions, further elaboration would be useful, but the information is there for the patient reader. On other occasions, the details are unnecessary, such as the exact tabulations of why each subject who left the experiment did so. These are minor issues that are partly an editor's responsibility, and more thorough editing would have helped the presentation.

Daniel P. Keating, *University of Minnesota*

Psychophysiology: Benchmark Papers in Animal Behavior, vol. 6

Edited by Stephen W. Porges and Michael G. H. Coles. New York: Halsted Press, 1976. Pp. 365. $25.00.

This book reproduces classical papers in psychophysiology (both human and animal) and "also includes several more recent papers that help bridge the gap between the theoretical views of the modern psychophysiologist and those of his predecessors" (from the preface). There are 24 papers divided into four parts. The editors comment briefly on each paper, giving some idea why they thought the paper important enough to be included. There is another preface by the editor of the series (Martin W. Schein), an index of authors cited in the papers, and a subject index.

Part I is entitled "Methodology: Measures and Measurement." However, the eight papers in this section report, in some cases, original discoveries of phenomena as well as concerns with methodology. The oldest study in the book is an abstract of a paper Caton presented to the British Medical Association in 1875, the earliest report of electrical activity in the brain. The papers by Féré in 1888 and by Tarchanoff in 1890 on the psychogalvanic reflex are translated and presented here, as is Berger's classical 1929 report on the electroencephalogram. In 1915, the monograph by Eppinger and Hess on the autonomic system was translated from German and published in English; part of the translation is reprinted here. The remaining three papers are by Walter and others on contingent negative variation (1964) and by Wilder (1931) and Wenger (1941) on the autonomic system.

Part II, on "Arousal Theory," contains five papers. One of them is, I am happy to say, the classic paper by Yerkes and Dodson in 1908. This paper is still worth reading, all the more so because it reports not only the Yerkes-Dodson law but another major finding. The latter was that the law interacted with the level of difficulty of the task, so much so that the law did not hold at all levels of difficulty. This paper must be one of the most widely cited in psychology, but it is still not cited widely enough; both findings of the paper continue to be

rediscovered. The other papers in this part are by Lindsley on behavior and the electroencephalogram (1952), Hebb on drives and the central nervous system (1955), Duffy on arousal (1957), and Malmo on anxiety and arousal (1957).

Part III, on "Orienting Reflex and Attention," consists of five papers on a variety of topics in this diffuse area. They are the papers by Sokolov on perception and conditioning (1963), Darrow on physiological reactions to stimuli (1929), Lacey on stress and activation (1967), Graham and Clifton on heart rate and orienting response (1966), and Obrist and others on cardiac/somatic relationships (1970).

Part IV has six papers under the heading "Emotion and Autonomic Conditioning." James's classic paper of 1884 on what became known as the James-Lange theory of the emotions is here and is still worth reading. So is Cannon's 1927 paper, which strongly attacked but did not destroy that theory. In fact, the third paper in this part, by Ax on physiological differentiation of fear and anger (1953), disputes one of Cannon's major points. The remaining papers are by Schachter and Singer on determinants of emotional state (1962), Miller on visceral and glandular conditioning (1969), and Katkin and Murray on conditioning of autonomic behavior (1968).

This book provides a valuable service in making available a number of important papers, several of them translated for the first time. My only criticism is that in the attempt to bridge 95 years of work (1875–1970), the book seems disjointed. It contains both history and fairly modern theory. It would be nice to have a volume on each of these topics, but since that hope may be unrealistic, I am grateful for this book.

C.P.D.

Eminent Contributors to Psychology
Vol. 2, A Bibliography of Secondary References

By Robert I. Watson. New York: Springer Publishing, 1976. Pp. 1,158. $80.00.

Watson's first volume (on primary references), published in 1974 and reviewed in this journal (1975, 88:707–708), was based on over 500 contributors to psychology, rated for eminence. The second volume now provides selected secondary references for these eminent contributors — references to them and their work by others. The methods used to search for and to select the secondary references are described in the introduction. There is a name index of the eminent contributors as cited by other eminents. Finally, there is a brief addendum that covers some of the literature through 1973.

As its size and price indicate, this volume represents an unbelievably enormous amount of work. It includes about 55,000 references. There are over 1,200 references for Freud. But Watson points out that there is a published bibliography of over 90,000 references to psychoanalysis through 1970. And what reference to psychoanalysis does not cite Freud directly or indirectly? Similarly, Watson selected and included over 300 references to Rorschach, but he notes that well over 4,000 references to Rorschach are cited through Buros' *Seventh Mental Measurements Yearbook*. Thus, although Watson did have to locate secondary references (not easy for some of the eminents), the really staggering

job was winnowing and selecting among the huge mass of material. Also, in a book of this size, the task of verifying and cross-checking for accuracy was a big job in itself.

In this context, it is pleasant to report that Watson includes book reviews among the secondary references and praises them as good sources of information about a contributor. A book review is often the only place where a man's major work is directly evaluated by a contemporary.

In my review of the first volume, I expressed a hope that it would be bought not only by libraries but also by individual psychologists. This seemed a reasonable hope, since the book cost $27.00. Although the second volume should also be bought by libraries, its price will deter some individuals. (Both volumes can be purchased as a set for $95.00.) This is too bad, although the price is not out of line for the size of the book. The second volume shows how the work of a man is imbedded in the minds of his contemporaries and successors, and in this respect it provides important information that cannot be obtained from the first volume. Unless others read and cite what a man has written, he is not likely to achieve eminence or be remembered.

In producing these two volumes, Watson has made an outstanding contribution to the study of psychology and its history. I hope his work comes to be widely cited.

C.P.D.

Patterns of Psychological Thought: Readings in Contemporary and Classical Texts

Edited by James R. Averill. New York: Halsted Press, 1976. Pp. 603. $22.50.

To quote from the preface, "this anthology differs from previous ones in two ways. First, it attempts to sample relatively few historical figures (10), each at sufficient length to allow the student to become involved in the author's work. Second, each historical excerpt is paired with the writings of a contemporary theorist. The purpose is . . . to show how both have something to say about an issue of contemporary concern."

Each of the 11 sections of the book has two chapters. In the first section, there is a long introductory chapter by Averill and a selection from Thomas Kuhn on dogma in scientific research. Averill has also written a brief introduction to each of the 21 selections from other authors.

The other 10 pairs of authors are: Plato and Lawrence Kohlberg on truths and values; Aristotle and Charles Taylor on types of psychological explanation; Plotinus and Charles Tart on states of consciousness; Augustine and Theodore Sarbin with Nathan Adler on the limits of reason and will; Aquinas and Magda Arnold on the meaning of emotion; Descartes and Noam Chomsky on the mental and the physical; Hume and Gordon Allport on empiricism and the illusion of knowledge; Kant and Konrad Lorenz on the problem of knowledge reformulated; Darwin and B. F. Skinner on the evolution of behavior; and Marx and Alexander Luria on the institutionalization of behavior. There is a short epilogue, a name index, and a subject index.

It is good to have a source book with relatively long selections from each author. In many cases, one feels he is getting the 'flavor' of an author's thinking

in a way that is not possible from the usual short selection. The problem, of course, is that relatively few writers can be covered. Another problem with long selections, one the editor can't do much about, is that the selections will vary in interest and readability to different readers. I found Kuhn, Augustine, and Darwin interesting, Plotinus and Marx turgid, Aristotle dull, and so on. The editor's introductions are helpful summaries for content, but the widely varying prose styles must simply be endured. Despite this, the book is a worthwhile attempt to relate historical and contemporary views of persistent problems in psychology.

C.P.D.

A History of Clinical Psychology

By John M. Reisman. New York: Halsted Press, 1976. Pp. 420. $15.95.

This book is an enlarged edition of the author's *The Development of Clinical Psychology,* published in 1966. In the current work, the first chapter is devoted to the eighteenth- and nineteenth-century precursors (individuals and events) of clinical psychology. Each of the succeeding eight chapters is devoted to a decade, 1890–1899 through 1960–1969. References follow each chapter. There is a combined name and subject index.

The organization within each of chapters 2 through 9 is the same. First, a few pages are devoted to general American history and culture of the particular decade being dealt with. This is a laudable attempt to provide a broader context for the developments in clinical psychology in that decade. Then the bulk of the chapter is organized under these headings, in this order: on normal personality functioning, diagnostic techniques, diagnostic formulations, treatment formulations, and professional development. In general, these sections respectively cover personality theory, tests, abnormal psychology, methods of treatment, and the growth of the profession of psychology as a whole and of clinical psychology in particular.

I found little to criticize in this book, but one detail bothered me. For some well-known psychologists the year of birth (and of death, if appropriate) is given in the text, but for many others the dates are not given. Why some but not others? Also, for Abraham Maslow and George Kelly, the year of birth but not the year of death is given. On the whole, however, this is a thorough and informative book.

C.P.D.

Motivation: An Experimental Approach

By Eva Dreikurs Ferguson. New York: Holt, Rinehart and Winston, 1976. Pp. 453.

I approached the task of reviewing this book with considerable curiosity and with the feeling of 'coming back' as a visitor to a place once called home. Ten years ago, as a graduate student at the University of Texas, I was surrounded by active debates on the issues of motivation and learning, the debaters in residence being such luminaries as Kenneth Spence, John Capaldi, and

John Theios. Now, as a social psychologist, for whom the issues of motivation are often more implicit than explicit, I was eager to regain perspective on the field. Eva Ferguson's *Motivation: An Experimental Approach* promised to provide this reacquaintance.

The aims of the text, as stated in the preface, include the integration "of a wide area of psychology . . . too long fractionated" and the provision of an "overall perspective." To accomplish these aims, Ferguson has indeed covered a large number of topics.

In the first two chapters, she attempts to lay the groundwork by providing a general discussion of the construct of motivation and the methodological issues inherent in defining a construct and operationalizing variables. Unfortunately for some readers, Ferguson has chosen to gear her coverage more to empirical than theoretical issues. As a consequence of this choice, the broad background of theorizing in motivation is skimmed very briefly, and the interested reader is referred to sources such as Robert Bolles for the broader picture. Consistent with her stated goals, Ferguson goes into much greater detail on the nuts and bolts of conducting experiments, including a third chapter on such topics as using statistics and writing up an experiment.

The coverage then turns to substantive issues in motivation. Three long chapters are devoted to the issues of drive, activation, and arousal, and define what appears to be the major thrust of the volume. In pages alone, these three chapters constitute approximately a third of the total text. Many of the controversial issues in this area are covered in detail, and resolution is necessarily incomplete. Yet, one occasionally suspects that the balance between complexity and clarity has been tilted. While simplicity is often misleading, selectivity in presentation of issues and examples can minimize the confusion the novice may experience. Generally, Ferguson succeeds in this attempt, but there are some points of lapse.

A well-written chapter by Richard Musty deals with homeostatic drives and consummatory behavior, specifically hunger and thirst. This chapter excels, not only because of its clear organization and readable style but because it conveys some of the detective-story flavor of scientific investigation. This account of the search for the causes of food and water consumption cannot help but intrigue the reader.

Two additional chapters, well within the domain of traditional research on motivation, deal with drive, reward, and incentive, and with escape and avoidance behavior. Finally, the last three chapters of the text cover numerous areas of social motivation, including the need for achievement (with some reference to recent developments in attribution processes), aggression, love and attachment, and sexual behavior.

Combined with this substantive coverage of topics is an emphasis on actual experimentation. In addition to the chapter on designing experiments noted above, each chapter (with the exception of the last two chapters on social motivation) is accompanied by a set of experiments students can perform to highlight some of the issues discussed in the chapter. The inclusion of these experiments is an appealing feature. In each case, the experimental procedures are clearly explained and the apparatus kept to a minimum. Even in a course where actual experimentation was not feasible, I think, the author's careful explanation of procedures would help the student understand how a psychological experi-

ment is conducted. Summaries at the end of each chapter should also be welcomed by the student.

Given the relatively wide scope of the text and the author's recognition that not all topics can be covered, it may be unfair to suggest that not enough is included. Perhaps the question is not entirely one of scope, but one of balance. Admittedly, my bias here reflects my viewpoint as a social psychologist. For example, expectancy-value theories of motivation seem deemphasized in comparison to drive and arousal models. Other areas of social motivation are also given little coverage. Consistency models, for example, are barely mentioned. Finally, the numerous personality models and theories, while perhaps less experimental than one would wish, nonetheless would seem to deserve some place in a text that aims to be broadly integrative in its coverage.

Perhaps more serious in my mind, however, are the errors of commission as opposed to those of omission. The last three chapters, dealing with need for achievement and a variety of social motivations, are clearly below the quality of the preceding chapters. Indeed, they resemble a grab bag more than a systematic coverage, and the basis for selecting certain topics and ordering certain topics is quite unclear. Rather than being part of an integrated text, they appear instead to have been tacked on late in the planning and never totally linked to other material.

For example, in the chapter on achievement, the attribution and locus-of-control positions are not clearly distinguished. Curiously, the related material on fear of success is not included in this chapter but appears in a later chapter following a consideration of Machiavellianism.

I found the chapter on aggression similarly dissatisfying. Ferguson reviews some portions of the animal data in reasonable fashion, but as she moves on to review the human data, the discussion loses considerable rigor. Aggressiveness, as generally defined by the social psychologist to include an intent to do harm, is expanded to include outspoken students and assertive businesspeople. While the author acknowledges that assertiveness and aggressiveness may be different states, the confusion persists throughout the chapter. Further, one can question the use of Milgram's experiment on obedience as a prototype of aggressive behavior. Next comes a brief illustration of cognitive dissonance, its appearance here an example of the confusing sequences and combinations in the chapters on social motivation. Sex-role stereotypes, the Prisoner's Dilemma, and Alfred Adler also make brief appearances in this chapter. In summary, the chapters on social motivation are the least satisfying aspects of the book.

Finally, some comments on the style of the book are in order. While generally well written, the text does include numerous sections where the style becomes ponderous — far closer to academic journalese than is appropriate for an undergraduate audience. In addition, Ferguson makes two stylistic choices that annoyed this reader. First is the frequent use of the masculine pronoun when referring to the general case. This conspicuous pattern begins on the first page of the book, four times alone in the first paragraph on the "student who has had introductory psychology courses." Surely, everyone realizes that women take psychology too. By now we have the right to expect that textbooks, guided by either author or editor, will be neutral in gender. A second, and perhaps less serious criticism, is the repeated use of *student* in the abstract. "Students should" and "students will" grated on my ear; the choice serves to detach the reader (who in most cases will be a student) rather than to involve.

In summary, I found my reacquaintance with motivation somewhat disappointing. The integration and overall perspective for which the author aimed were not, in my opinion, achieved. At the same time, much of the material covered is indeed fascinating and one can agree with the author that motivation is an "exciting area of inquiry," even if the overall perspective is still lacking.

Kay Deaux, *Purdue University*

Behavioral Foundations of System Development
By David Meister. New York: Wiley, 1976. Pp. 373. $21.95.

Behavioral Foundations of System Development is a rather unusual book. It sets its own style and, to a large extent, defines the world of behavioral research to fit its own outlook. Author David Meister would not think it a bit presumptuous to do so because, as the editor of the Wiley series in human factors, he has given it a mission to tell the world, and especially the design engineer, what human factors is. The book is a volume in that series. As a result of these circumstances, a significant portion of this review will deal with Meister's definition of the field versus the actual content of the book.

First, the book is not, as the title would suggest, a study of the behaviors that underlie the developing of systems, an area that sorely needs examination. Also, the basic material of the book, as the dust cover says, consists of a set of cases. But the cases are not case studies in system development. In fact, they are not cases at all in the traditional sense. It is a convenient way for Meister to identify single research studies. The stated purpose of the book is to determine what the behavioral literature, conceived broadly as personnel functioning, has to say that is meaningful for system development. The task, however, turns out to be considerably less grandiose, because Meister rules beforehand what portions of the literature can and cannot be usefully applied in system development. Research that focuses on the individual and does not use a complete function for the experimental task is considered to be molecular and abstract, and not applicable to system development without great caution. This restriction essentially eliminates most basic and applied psychological research and the complete area of human engineering — the so-called knobs-and-dials work that has to do with the interface of the individual with his or her work environment. The latter may come as a surprise to those who identify themselves as 'human engineers.' The literature that does qualify as meaningful for system development would use a system context, complete functions, and two sets of measurements — one set for the human in the system and another for the system itself. The author finds that these ideal conditions are difficult to meet.

Actually, the literature covered becomes even more restricted, because Meister says that he just does not have much interest in vehicles and aircraft and the general area of manual control. Other areas are eliminated by not being mentioned at all or only in passing. The broad range of sociotechnical systems — such as a ship and its crew, a health-care delivery system, a manufacturing operation, a building complex — come to mind in this category of areas left uncovered. What he does cover are many studies that have individuals or several individuals sit in front of electronic or electromechanical displays and push buttons and levers or call out their responses. The settings for these studies

are, in the main, analogues of air traffic control or air combat information centers.

The core of the book is the so-called cases, and it is the discussion of these cases that forms the author's behavioral foundations of system development. The cases are augmented by briefer summaries and mentions of related studies, especially more recent ones. The cases and the general references make up a very substantial body of literature in the general area of human performance. Government reports have been identified by their AD numbers so that they can be ordered from the National Technical Information Service or the Defense Documentation Center. There are 191 cases, of which 29 are from before 1960, 66 are from 1960 to 1964, 65 are from 1965 to 1969, and 31 are from the 1970s. Thus, it is obvious that they are primarily from the 1960s, that heyday of the engineer when defense and space dollars seemed unlimited, when the human-factors specialist could raise his voice, insist that he be listened to, and call for the study that would provide the empirical answer to a particular question of design. The result was a profusion of studies with much face validity and often not much else. There were also, embedded in program research, certain studies with only incidental face validity, studies in which the systematic investigation of the capability of the adult human to process information and make decisions was the crucial concern. These studies were very much person- or individual-oriented and formed the basis for the very popular orientation that today studies the human as an information-processing system. Both types of studies are included among the cases.

Meister's concept of a human in a system must be understood to appreciate what *Behavioral Foundations of System Development* is trying to say. A system is considered, quite conventionally, to be a collection of men and material with a purpose. Perhaps unconventionally, Meister asserts that a man/machine system is a biological system because it is controlled by an operator. The crucial point, which Meister takes great pains to emphasize, is the difference between inputs and outputs as distinguished from stimuli and responses. The former describe subsystem and system processes and the latter are said to describe internal human processes. The means by which inputs are transformed to stimuli and responses to outputs are the province of the researcher in human factors. The processes that interrelate stimuli and responses are the province of the psychologist, since they occur solely within the human.

To give an example of this two-stage operation at the input and output ends of the chain, let us look at what Meister does in the book with the concept of load The system input may be enemy aircraft or their representations on a radar scope, which can increase in frequency, number, or both; but load does not become a stimulus until the individual feels it as such. That is, it is not the frequency, number, or density of enemy aircraft that defines load, it is a condition within the individual that does. Therefore, in this example, human factors is the study of how changes in the environment become transformed into load for the individual and of how responses to that load, once induced, result in changes to the system's output.

The book, in addition to introductory and concluding chapters, is divided into five substantive chapters that represent Meister's first level of differentiating the human-factors literature. These are on the effects of input load on system performance, task characteristics, decision making, team functions, and feedback.

Each chapter is then broken down into several areas, each with a definition and some explanatory remarks. Within these second-level categories, there are many sets of conclusions followed by illustrative cases. Then there is a summary of conclusions followed by a section called developmental implications. It is rather disconcerting to come upon a conclusion in the middle of a narrative passage and to have to remind oneself that it pertains to the material to follow, not to that before. This occurs frequently because there seems to be at least one conclusion for each case, if not more (because each case may contribute to several conclusions).

The material in the book is very fragmented, and no effort has been made at integration. Problems in organizing the material arise because the same phenomena can be, and are, treated in several parts of the book. Second-level elements in the chapter on task characteristics, elements such as difficulty and complexity, turn out to be essentially the same as load inputs in the preceding chapter. Organizational characteristics are discussed primarily in the chapter on task characteristics rather than in the one on team functions, and they also appear as inputs to load in that chapter. Meister attributes these problems to the fact that the variables are of a higher order (superordinate) in a system context. The inability to integrate the findings is also attributed to the "macro nature" of the variables. These variables are said to interact with so many features that the principles described can only be understood in terms of the original experimental conditions.

Commonly accepted psychological concepts and practices appear to be used indiscriminately or interpreted differently. Items such as response requirements and payoff matrices that are traditionally placed at the output end of the behavioral chain are treated as inputs because they contribute to load. System-status inputs to a monitoring system are treated as feedback because they are experienced that way by the individual. The term 'interaction' is everywhere evident, and it apparently means any effect between two variables that is different from their individual effects, whether they be additive or multiplicative. The inverted U of Hebb's cue function is applied directly to performance rather than to cue utilization. It is asserted that information measured in Shannon 'bits' is inappropriate in the system context because bits can only be applied to very discrete behaviors. Instances such as these could possibly be explained by Meister's contention that human factors is a "distinctive discipline" that has its own rationale "and makes use of psychological principles, but in their application . . . goes well beyond their source."

The summaries of conclusions in the five substantive chapters read well and provide many important generalizations. The conclusions are mostly about the behavior and performance of individuals and groups, not systems. A few are patently ad hoc, and although they are based on the illustrative cases, they are inappropriate as generalizations in the light of the broader scientific literature. The sections on developmental implications are much less consistent in quality. They cover a good range and seem quite appropriate in the chapters on input load, decision making, and feedback. In the lengthy chapter on task characteristics, though, aside from advice about organizations and teams, the only implication seems to be Keep Things Simple. The section in the chapter on team functions consists primarily of training recommendations from one source. Moreover, the guidance provided by the summaries of conclusions and the sec-

tions on implications for design is all qualitative, and there is no way for the design engineer to apply it directly and unambiguously to specific cases.

The problems and shortcomings of the book that are mentioned above are recognized by Meister in his concluding chapter as characteristics of the discipline. He explains their roots and how they must be overcome. The elements that would seem to have had the greatest weight in creating the present situation are the simplistic demands made of human-factors specialists by employers naive in the behavioral sciences and the penchant for the human-factors specialists to be, in turn, overly practical, pragmatic, and atheoretic when they are frequently dealing with constructs that properly demand systematic elaboration.

James K. Arima, *Naval Postgraduate School, Monterey, California*

The Thinking Computer: Man inside Matter

By Bertram Raphael. San Francisco: W. H. Freeman, 1976. Pp. 322. $12.95.

Artificial intelligence can no longer be called a new field of endeavor. It is an established discipline. Yet it is still surrounded by the controversy and misunderstandings that normally attend a new area. Preposterous and unfulfilled prophecies by some of its practitioners and outrageous and unscientific criticisms by some of its detractors have contributed to the atmosphere of antagonism. Perhaps a more important factor, though, has been that until recently the body of knowledge accumulated by researchers in the field was seldom accessible except through the numerous computer programs available only to the 'in' group of practitioners or through fairly technical articles. In the past five years, however, several textbooks on the field of artificial intelligence have appeared. *The Thinking Computer* by Bertram Raphael seems unquestionably the best of the new textbooks, especially for psychologists and for students without a background in computer science.

The book is sophisticated and technically accurate, yet it is written with a style and a wit that make it delightful reading. Raphael displays a refreshing irreverence toward the clichés and dogma of information processing. In discussing "information and knowledge," Raphael says, for example, "the word 'information' once had a perfectly good, generally understood meaning — a meaning that most of you probably still have in mind. Then the word was ruined by communication scientists, who gave it a very precise but somewhat different meaning for their own purposes. . . . We shall use the word 'knowledge' to refer to what used to be called 'information.' . . . However, we shall be careful not to give 'knowledge' a precise technical definition, and thereby ruin the word for future generations" (p. 47). Raphael also demonstrates that 'one picture is not worth 1,000 words,' but rather "about 10,000 words." Throughout the book he has included considerable interesting trivia and puzzles that, while in no way central to the topic, will stimulate and amuse anyone with a bent toward artificial intelligence. In a sense, these convey the flavor of the field as well as any part of the book.

The book begins with an excellent, concise discussion of computing hardware, programming languages, and their limitations. Then the reader is exposed to issues that have arisen in the characterization of problems for computer

solution and to the importance of the representation chosen for the solution process. Both simple abstract examples (e.g., the mutilated chessboard) and real problems for which solutions have been programmed are presented. For example, Samuel's elegant solution for an efficient representation of a checkerboard is described in great detail. The reader is also introduced to representation with strings, list structures, and predicate calculus.

Raphael next turns to a discussion of the theory of searching for solutions to problems. After an informal introduction and several examples, he acquaints the reader with the formal theory of solution trees and tree search. The organization of this section on search algorithms is very similar to that in Nilsson's textbook, but more readable for the untutored. These chapters should be required reading for any cognitive psychologists interested in problem solving.

Chapter 4, strangely mislabeled "Problem Solving Methods," deals with pattern recognition and theorem proving by computer. There is no way in a general text that one could do justice to the massive literature on pattern recognition, and Raphael in this chapter only briefly summarizes the general theory of pattern recognition as developed by researchers in the field of artificial intelligence. Later in the text, he does present a detailed discussion of picture-processing techniques. While this section will be valuable for students, psychologists interested in visual perception will probably have encountered all this material elsewhere. The part of the chapter devoted to theorem proving, however, contains the most intelligible introduction to theorem proving by computer that I have seen. For propositional calculus, both Wang's algorithm and propositional resolution are illustrated, while for predicate calculus, the resolution approach is clearly explained.

These sections are followed by excellent chapters on general problem-solving systems and language-understanding systems. Having instructed the reader in theorem-proving methods first, Raphael can use fairly sophisticated examples involving the representation of problems in predicate calculus. The section on general problem solving goes far beyond the usual few words devoted to the matter and into the more sophisticated general problem-solving systems developed recently. Raphael's discussion of machine learning in problem solving is too brief and simplistic to be very valuable, however. The chapter on natural-language processing by computer will be an excellent introduction to the topic for naive students, but again most cognitive psychologists will already be quite familiar with the material.

The final two chapters concern robotics and other applied uses of artificial intelligence. Many psychologists are not very familiar with robotics and may be amazed to discover some of the advances that have been made, though I suspect one must at some time have tried to program some of these tasks in order to truly appreciate the advances.

Overall, this book would be an excellent text to assign as part of an advanced course in cognitive psychology or a course on the methodology of modeling and computer application. In addition, most of the book, especially its coverage of problem representation, search, theorem proving, and robotics, would be valuable reading for any cognitive psychologist.

L. Rowell Huesmann, *University of Illinois at Chicago Circle*

ANNOUNCEMENT

American Association for the Advancement of Science: 1977/78 Chautauqua-type short courses for college teachers

Come October the 1977/78 series of NSF Chautauqua-type short courses for college teachers will offer the opportunity for some 3,500 college teachers to participate. Approximately 50 short courses will be offered in the 1977/78 academic year at field centers throughout the United States. The program is designed to enable college teachers to keep abreast of advances in a variety of field of science and to help them incorporate these advances in their teaching. The program is conducted by the AAAS with support from the National Science Foundation.

Typical topics during the 1976/77 series were: Origins of Life, Ethical Issues in the Life Sciences, Solar Energy, Psychology of Women, Biofeedback, Fundament Particles: Quarks to Quasars, Genetics and Society, Plate Tectonics, Privacy, Population, and Microcomputers Applied to Science Education.

Participants meet in groups of 25 for two-day sessions in the fall and early spring. During the interim, participants are able to work individually or in teams on problems or projects related to the courses.

Further information is available from the AAAS. For course titles and course directors, locations, schedules, eligibility, and application forms, please write to

American Association for the Advancement of Science
Office of Science Education, Box G9
1776 Massachusetts Avenue, N.W.
Washington, D.C. 20036

REPORT

The Seventy-third Annual Meeting of the Society of Experimental Psychologists

The seventy-third annual meeting of the Society of Experimental Psychologists was held at Yale University, New Haven, Connecticut, on April 1 and 2, 1977. Wendell Garner and Alvin Liberman, cochairmen of the society for the year, presided at the business meeting and at sessions for the presentation of scientific papers.

The members and fellows attending the meeting were D. Blough, C. Buxton, E. Carterette, W. Estes, F. Finger, R. Gagne, W. Garner, E. Gibson, F. Graham, D. Grant, D. Green, R. Grice, I. Hirsh, J. Hochberg, D. Hurvich, L. Hurvich, W. Hunt, T. Kendler, H. Kendler, G. Kimble, J. Lacey, H. Leibowitz, A. Liberman, D. Lindsley, F. Logan, D. Luce, D. Meyer, N. Miller, B. Murdock, C. Pfaffmann, R. Rescorla, B. Rosner, G. Sperling, P. Teitelbaum, W. Torgerson, E. Tulving, J. Volkmann, A. Wagner, and D. Wickens. The membership elected Mathew Alpern, Beatrice Lacey, Fergus I. M. Craik, David Premack, Janet T. Spence, and David Zeaman as new members. These new members and the deaths of James Olds and Hans-Lukas Teuber bring the combined membership to 81 members and 58 fellows.

Reports of research were presented in scientific sessions on both days of the meeting. The Warren Medal for 1977 was awarded to Eleanor J. Gibson "for her extensive experimental and theoretical contributions to psychology in the fields of perceptual development, verbal learning, and reading."

The society accepted the invitation of UCLA to meet in Los Angeles, California, in 1978. Edward Carterette and Donald Lindsley were elected cochairmen for 1977/78.

Gregory A. Kimble, Secretary; *Duke University*

Contents

Louis G. Lippman and Richard W. Thompson. The effect of shock on the exploratory behavior of rats in a complex maze 363

Henry M. Halff. The role of opportunities for recall in learning to retrieve 383

Charles F. Hinderliter and David C. Riccio. Long-term effects of prior experience in attenuating amnesia 407

James W. Pellegrino and Judith Petrich. List discrimination during transfer in free recall 419

Robert E. Hicks, George W. Miller, Gerald Gaes, and Karen Bierman. Concurrent processing demands and the experience of time-in-passing 431

James H. Neely. The effects of visual and verbal satiation on a lexical decision task 447

Thomas R. Herzog and Daniel J. Weintraub. Preserving the kinesthetic aftereffect: Alternating inducing blocks day by day 461

James W. Aldridge and Michael T. Farrell. Long-term recency effects in free recall 475

Donald H. Kausler, David M. Kleim, and Robert M. Yadrick. False recognition of synonyms of right and wrong items in multiple-item recognition learning 481

Douglas L. Hintzman and Leonard D. Stern. Failure to confirm Elmes, Greener, and Wilkinson's finding on the spacing effect 489

John C. Thomas, James L. Fozard, and Nancy C. Waugh. Age-related differences in naming latency 499

James D. Evans. On the inconstant effects of study instructions 511

Lois M. Freed and William J. Freed. Effects of schedules of reinforcement on hypothesis-refining behavior 517

Robert Sommer. A note on 'graphonyms' 527

H. D. Kimmel. *Notes from "Pavlov's Wednesdays": Partial reinforcement as a test of mobility* 529

Book reviews (Keith J. Holyoak, Eugene B. Zechmeister, Randolph Blake, Mark S. Mayzner, Aryeh Routtenberg) 533

AMERICAN JOURNAL OF PSYCHOLOGY
September 1977, Vol. 90, No. 3, pp. 363-382

The effect of shock on the exploratory behavior of rats in a complex maze

Louis G. Lippman and Richard W. Thompson
Western Washington University

As replication and extension of prior work, two experiments investigated the activity and exploratory behavior of rats during 10–16 days of testing in a complex environment — a +-maze with black, white, striped and checkered arms. Two measures were taken: frequency of entrance into each arm and a sampling of each animal's location within each minute of each daily 3-min trial. Results of both experiments showed consistent trends in the animals' topographies of exploratory activity and configuration of arm preference. Shock-induced fear reinstated the preferences for low-intensity and low-complexity arms exhibited at the outset of testing, thus providing evidence for the contribution of fear to exploratory activity.

Although exploratory behavior of the rat has been the subject of much investigation and theorizing (Berlyne, 1960; Bronstein, 1972; DeNelsky and Denenberg, 1967; Fiske and Maddi, 1961; Halliday, 1967; Lester, 1967; Montgomery, 1955; Russell, 1973), the majority of these investigations have used a simple open-field or Y- or T-maze situation. In a recent series of experiments, Thompson and Lippman (1972, 1975) have investigated the activity and exploratory behavior of rats using a +-maze (a Greek cross) with black, white, striped and checkered arms — a more complex testing environment. In these experiments, animals were given a single 3-min test on each of two consecutive days. It was found, using the frequency of an animal's entering an arm of the maze and a sampling of the animal's location every 10 sec as measures, that repeated testing and habituation to handling tended to increase activity. Although the animals initially showed a strong preference for the black arm of the maze, avoidance of the white arm, and only intermediate preferences for the striped and checkered arms, repeated testing and habituation to handling equalized the animals' preference for the other arms of the maze. Thompson and Lippman hypothesized that repeated testing and habituation to handling decrease the animals' fear, which results in the changes in observed behavior.

The present experiments were designed to extend this series of experiments and to assess more directly the role of fear in the exploratory behavior in rats.

EXPERIMENT I

Experiment I extended the testing days from 2 to 12. It was expected that with repeated testing, the animals' fear would be reduced and their activity would increase, as would their preferences for arms of the maze other than the black one. To provide a more fine-grain analysis of behavior, records of each daily trial were partitioned as three consecutive 1-min periods. Further, to directly test the role of fear in exploratory behavior, half of the animals were subjected to electric shock after their activity and exploratory behavior had stabilized. (Electric shock has frequently been employed to induce fear in investigations of exploratory behavior in mazes; see Russell, 1973.) After two days with shock (days 13 and 14), the animals were given two recovery days in the maze without shock (days 15 and 16). The other half of the animals were treated in an identical fashion except that they were never shocked. All animals received 16 consecutive days of testing in the maze.

METHOD

Subjects

The subjects were 10 male albino rats (obtained from Sprague-Dawley, Madison, Wisconsin) housed 5 to a cage with ad lib food and water throughout the experiment. The rats were approximately 90 days of age at the beginning of the experiment.

Apparatus

The animals were tested in a maze identical to that described by Thompson and Lippman (1972). Its arms were 33 cm long, 25 cm wide, and 48 cm high. The floor of the maze was painted flat gray, and the north and south arms were painted flat black and flat white respectively. The west arm was painted with 2.5 cm black-and-white checks, and the east arm was painted with alternating 2.5-cm wide black and white stripes at a 45-deg angle. The maze was illuminated by two fluorescent lamps suspended 210 cm above the floor. The light falling on the floor in the center of the maze was 52 ftc. Measured as reflected from the walls of each arm, it was for the white arm, 16 ftc; black, 2 ftc; checkered and striped, 7 ftc. These measurements were taken with a Weston light meter (model 614).

The apparatus for shocking the rats was in a different room. It was a box 42 cm square and 40 cm high painted flat gray. The floor was of brass rods .5 cm in diameter placed 3 cm center to center. The grid floor was connected to a

C. J. Applegate stimulator through a Grason-Stadler grid scrambler (model E1064 GSP) and an electronic timer to provide for a 1.0mA shock of 3.0 sec.

Procedure

Each animal was given 10 days of handling, marked by painting its tail with dye, and then tested for three consecutive 1-min periods on 16 consecutive days in the maze. The handling was identical to that described by Lippman, Galosy, and Thompson (1970). Briefly, it consisted of removing the rats' home cage from the cage rack, transferring each rat to an identical holding cage, returning each rat to the home cage, and the cage to the cage rack. This procedure was repeated five times each day for 10 consecutive days, each animal being handled a total of 100 times. Testing consisted of placing each animal in the center of the maze facing one of the arms of the maze for a 3-min trial. The arm of the maze the rat faced was systematically rotated so that each rat faced each arm equally often. Two measures were taken: the animal's *frequency* of entering an arm of the maze during each of the three 1-min periods of each daily trial and a check on the animal's *location* every 10 sec throughout each daily session. Location within an arm of the maze was defined as the rat's head and two forelegs being in the arm. On days 13 and 14 only, the animals were placed in the gray shocking box for 23 sec just before the daily run in the maze. Half the rats received a single 3-sec shock of 1 mA 10 sec after being put in the box; the other half did not. All subjects were tested as before in the maze immediately after removal from the shocking box. On days 15 and 16 testing was conducted as on days 1–12.

RESULTS

The data for this and the following experiment underwent analysis of variance and only effects significant at $p < .05$ are reported. Significant interactions underwent analysis of variance and/or Duncan's multiple-range tests.

Replication (days 1–2)

Days 1 and 2 of testing were essentially a replication of the study by Thompson and Lippman (1972). Analysis of the data on *frequency* with factors of days, 1-min periods within days, and arms of the maze revealed significant effects due to arms and the interaction of days and arms [both $Fs(3, 27) \geq 3.10$]. Further analysis revealed that the black arm of the maze was entered more frequently than any other arm on day 1, but on day 2 it was entered more frequently than only the checkered and white ones. These results are consistent with those reported by Thompson and Lippman (1972). Similar analysis of the check on the animal's *location* every 10 sec revealed an effect only of arms [$F(3, 27) = 75.68$], indicating that the animals were more likely to be found in the black arm than any other.

Extended testing (days 1–12)

Analysis of the data on the animals' *frequency* of entering an arm of the maze with factors of days, periods, and arms of the maze yielded significant main effects of periods and arms and interactions of days and arms [$F(33, 297) = 2.75$] and days and periods [$F(22, 198) = 4.08$].

The interaction of days by periods is shown at the top of Figure 1. As is apparent there, the subjects were equally active throughout the three 1-min periods initially (nonsignificant interactions of days by periods on

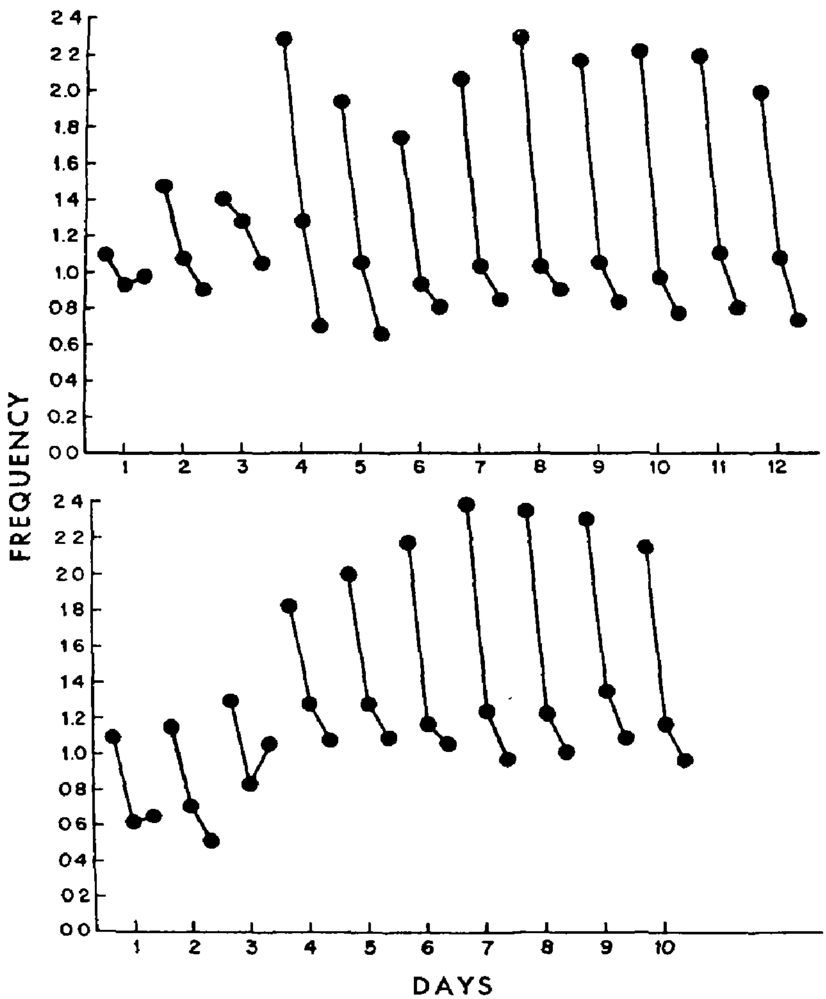


Figure 1. Frequency during each 1-min period over initial testing days (the interaction of days by periods): top, with the four arms combined, Experiment I; bottom, with arms and preplacements combined, Experiment II

test days 1 and 2), but with continued testing the exploratory activity became concentrated in the first 1-min period of the daily 3-min trial.

As can be seen at the left of Figure 2, the interaction of days by arms arose from the initial differences in arms entered on days 1 and 2 that then trended toward uniformity with continued testing. This trend can be seen as an increased preference first for the striped, then checkered, and finally white arms of the maze, although the black one still was entered more often than any other arm throughout the test days.

Similar analysis of the check on the animals' *locations* every 10 sec revealed a significant effect of arms [$F(3, 27) = 69.60$] and interactions of days by arms [$F(33, 297) = 5.93$] and periods by arms [$F(6, 54) = 5.96$]. The interaction of days by arms is presented at the left of Figure 3. As can be seen there, the changes in exploratory behavior seen in the data on frequency are also found in the data on location, which is independent of activity. Although the rat was most likely to be found in the black arm over all days of testing, the likelihood of being found in the other arms, especially the white one, increased with repeated testing. Examination of this plot (as well as the left panel of Figure 2) shows that the order of

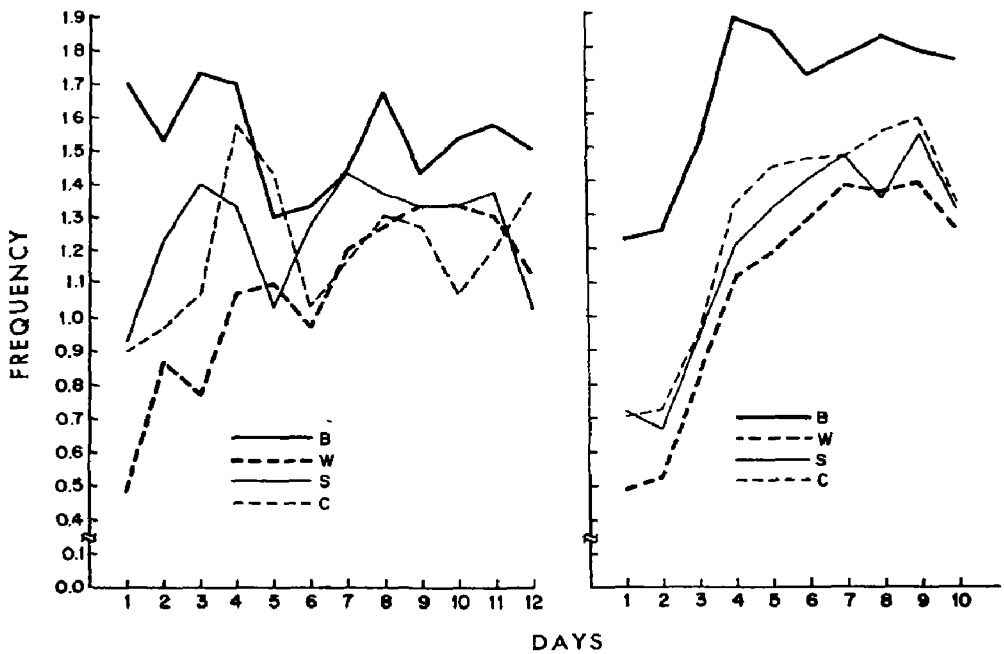


Figure 2. Frequency over initial testing days (the interaction of days by arms): left, with 1-min periods combined, Experiment I; right, with periods and pre-placements combined, Experiment II

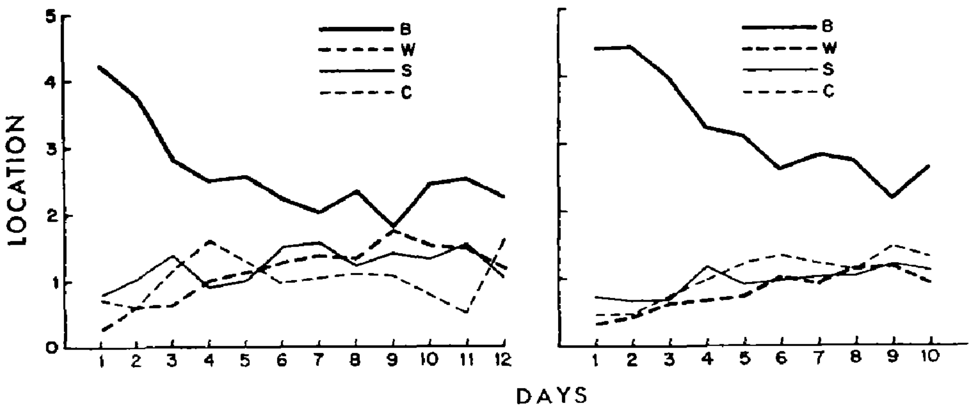


Figure 3. Location over initial testing days (the interaction of days by arms): left, with 1-min periods combined, Experiment I; right, with periods and preplacements combined, Experiment II

exploration and preference for arms of the maze is first for the black arm, then the striped, checkered, and finally white arms.

The interaction of periods by arms is presented in Table 1. Examination of this table reveals that an animal was equally likely to be found in either the striped or checkered arm throughout the testing period, decreasingly likely to be found in the white one, and increasingly likely to be found in the black one over the three 1-min periods. The animal, upon placement in the apparatus, appears to explore all compartments about

Table 1. Location (mean number of times rats were observed in a given arm of the maze at checks taken every 10 sec) during each 1-min period, combining both testing days 1–12 and 1–10 for the respective experiments and preplacements in the latter experiment (the interaction of periods by arms); Experiments I and II

	Arm of the maze			
	Black	White	Striped	Checkered
Experiment I				
First period	2.31	1.42	1.18	1.09
Second period	2 62	1 06	1.36	.97
Third period	2.95	.89	1.13	1.03
Experiment II				
First period	3.05	.84	1.06	1.06
Second period	3.28	.75	91	1.07
Third period	3.32	.79	.90	1.00

equally and eventually settles into the arm of choice, the black one, and avoids the white one.

Shock (days 13–14) and recovery (days 15–16)

The data on *frequency* were analyzed with factors of shock, days (within shock and nonshock conditions), 1-min periods within each daily trial, and arms of the maze. Results indicated significant main effects of shock, periods, and arms, and interactions of shock by periods and shock by periods by arms [$F(6, 48) = 2.70$]. This triple interaction, presented in Table 2, was further analyzed by analysis of variance for each period with factors of shock and arms. Analysis of the data on the first 1-min period revealed that for nonshocked subjects there was no difference in frequency of arms entered among the four arms, but for shocked subjects the black arm was entered most frequently, there being no differences among the other arms. During the second period, the analysis revealed that for both shocked and nonshocked groups the black arm was entered most frequently. During the third period, the shocked subjects were less active than the nonshocked subjects and both groups entered the black arm more frequently and the white arm less frequently than the other arms.

The effects of shock were most striking in the first period, where shock had the effect of reinstating a strong preference for the black arm similar to that observed on the first day of testing. Shock greatly reduced activity, this reduction being most observable in the first and third periods. There was no recovery of activity during the two days (15 and 16) after shock.

A similar analysis of the data on *location* revealed a significant main

Table 2. Frequency (mean number of arms entered) during each 1-min period, combining testing days 13–16 (the interaction of shock by periods by arms); Experiment I

	Arm of the maze			
	Black	White	Striped	Checkered
First period				
Nonshocked	2.40	2.00	2.00	2.30
Shocked	2.25	.90	1.35	1.05
Second period				
Nonshocked	1.15	.65	.70	.85
Shocked	.85	.55	.80	.50
Third period				
Nonshocked	1.10	.55	.95	.90
Shocked	.70	.30	.30	.50

effect of arms and a significant interaction of shock by arms [$F(3, 24) = 4.91$]. Further analysis of this interaction revealed that nonshocked subjects were more likely to be found in the black arm and were equally likely to be found in the other three arms. Shocked subjects were most likely to be found in the black arm too; however, they were also more likely to be found in the striped than either the checkered or white arm. This pattern of results indicates that shock tended to reinstate the preferences shown at the beginning of the experiment: a strong preference for the black arm and an avoidance of the highest intensities and complexities. There were no changes observed on the two days after shock.

DISCUSSION

The results of Experiment I are consistent with those reported earlier by Thompson and Lippman (1972). Rats given an opportunity to explore a maze with arms varying in intensity and complexity showed a preference — both in terms of data on frequency of entrance and on location every 10 sec — for the black arm and an avoidance of higher intensities and complexities. With repeated exposure to the maze, the animals explored the other arms more and were more likely to be found in other arms, the black arm remaining most preferred, followed by the striped and checkered ones, with the white one being the least preferred. By the end of 12 days of testing, the animals were visiting all arms of the maze about equally (at least at the outset of a testing session) and equally likely to be found in any of them.

The topography of exploratory behavior within each daily trial also changes. During the initial testing, the animals were about equally active throughout the three 1-min periods. Starting on day 4, however, the rats exhibited most of their exploratory behavior during the first 1-min period in the maze, there being a decrease during the next two 1-min periods. In terms of the data on location, after exploring the maze during the first period, the animals tended to 'settle in' to the arm of choice, the black one, and remain there for the rest of the trial.

Subjecting the animals to shock caused on days 13–14 a decrease in activity and reinstated the animals' preference for the arms of lowest intensity and complexity. These are also the same arms that were the most familiar, having been explored more on days 1–12. The results obtained are consistent with the hypothesis that fear will reduce exploration and increase the animals' tendency to approach the most familiar of stimuli, especially those of low intensity and complexity (Aitkin and Sheldon, 1970; Halliday, 1967; Haywood and Wachs, 1967; Lester, 1967; Sheldon,

1968; Russell, 1973). Failure to find changes during recovery (days 15–16) may be because the reinstated fear had failed to dissipate in time to permit detection.

EXPERIMENT II

Experiment II was essentially a replication of Experiment I with many more animals and certain important modifications. Whereas in Experiment I the rats had been given considerable handling before testing, in Experiment II they were not. This permitted the assessment of changes in exploratory behavior in animals that were initially more fearful. In addition, just before each daily (days 1–16) trial in the maze, half the animals were preplaced in a box dissimilar to any of the arms of the maze and half were preplaced in a box similar to the preferred black arm of the maze that was to follow. On day 11, half of each of these groups were given a single shock trial in the box in which they had been preplaced on trials 1–10. Recovery from shock was extended to five days (days 12–16). These modifications permitted evaluation of the effects of additional exposure to stimuli either similar or dissimilar to those of the maze on the development of exploratory behavior in the maze. By shocking animals in a box that was either similar or dissimilar to the preferred arm of the maze, it was possible to examine any differential changes in the topography of preference and exploratory behavior as it developed over the extended recovery period.

METHOD

Subjects

The subjects were 44 male albino rats (Sprague-Dawley, Madison, Wisconsin) approximately 85 days of age at the beginning of the experiment. The rats were housed 5 to a cage and had ad lib food and water throughout the experiment.

Apparatus

The same maze used in Experiment I was used for testing. There were also two preplacement boxes 32 cm long, 25 cm wide, and 48 cm high, with grid floors of brass rods .5 cm in diameter, 3 cm center to center. One box was painted flat gray and the light falling on the floor measured 25 ftc. The other box was painted flat black and the light falling on the floor measured 22 ftc. Both boxes were connected through a Grason-Stadler grid scrambler and electronic timer to a C. J. Applegate stimulator which could provide shock of 1.0 mA. The grid scrambler (but not the shock) was on throughout the experimental session. An electronic timer, which provided an audible click, ran throughout the experimental session and permitted the experimenter to time the various observation intervals.

Procedure

Each subject was marked for identification by dyeing its tail five days before testing; they received no other handling before testing. All animals were tested on 16 consecutive days. Each daily trial was again divided into three 1-min periods. Testing consisted of taking each animal to the experimental room, preplacing it for 30 sec in the gray or black box, removing it from the box and holding it for 20 sec, then placing it in the center of the maze. The arm the animal faced was systematically varied from trial to trial so that each subject faced each arm equally often. The design of the experiment provided that half of the subjects be preplaced in the gray box and the other half in the black box. Each of these two groups was further divided so that half of each received a single 2-sec shock of 1.0 mA 10 sec after being preplaced in the same box as before on day 11, and that day only. Two measures were taken: *frequency* of entrance into each arm of the maze during each 1-min period, and the animal's *location* every 10 sec throughout each daily trial.

RESULTS

Replication (days 1–2)

The data for days 1–2 provide a replication of Thompson and Lippman's study (1972) and of Experiment I. Analysis of the data on *frequency* with factors of preplacement, days, periods, and arms of the maze yielded results consistent with those obtained previously. There was a decrease in activity across the three periods, most dramatically from the first 1-min period each day to the second, especially for the black arm. Overall, the black arm of the maze was entered more frequently and the white one entered less frequently than any other arms. Of special interest is the interaction of preplacement by periods by arms [$F(6, 252) = 2.24$]. Animals preplaced in the black box showed less extreme differences in frequency of entrance into all arms, especially black versus white, than did those preplaced in the gray box, this difference being most apparent in the first period. Rats preplaced in the black box ordered their entrances into the arms of the maze: black, striped, checkered, and white. Rats preplaced in the gray box ordered their entrances: black, checkered, striped, and white. These differences were most apparent in the first period on each day.

A similar analysis of the data on *location* for days 1–2 revealed an interaction of periods by arms not seen in Experiment I [$F(6, 252) = 3.52$]. Examination of this interaction revealed that during the first 1-min period, the animal was most likely to be found in the black arm and least likely to be found in the white and checkered arms. Over the next two periods, the likelihood of its being found in the black arm increased, while the likelihoods of its being in the other arms decreased. The absence of

such an interaction in Experiment I may be because the animals in Experiment II were more emotional as a result of the preplacements before testing and the absence of an habituation to handling (which has been suggested by Thompson and Lippman to reduce emotionality and enhance exploratory behavior).

The absence of an effect of days or its appearance in an interaction in both these analyses is different from the results obtained in Experiment I and those obtained by Thompson and Lippman (1972). The deletion of habituation to handling and the addition of the preplacements may have resulted in greater emotionality during testing, yielding the resultant low level of activity and enhanced avoidance of arms of high intensity and complexity over days 1–2. These differences are readily apparent by comparing the two parts of Figure 1 for the frequency measure on days 1–2, and the two parts of Figure 2 for the location measure on the same days.

Extended testing (days 1–10)

Analysis of the data on *frequency* with factors of preplacement, periods, days, and arms revealed significant main effects of days, periods, and arms, and interactions of days by periods, days by arms, periods by arms, days by periods by arms [$F(54, 2{,}268) = 1.53$] and preplacement by periods by arms [$F(6, 252) = 2.83$].

Despite its presence in higher-order interactions, the effect of days by periods shown at the bottom of Figure 1 is of interest in comparison to the data of Experiment I shown at the top of Figure 1. Although the overall appearance of these two plots is similar, two differences should be noted. First, the animals in Experiment II were initially less active than those in Experiment I. Second, whereas there were no changes in activity in Experiment I, activity increased across days in Experiment II. These differences further support the interpretation that the animals in Experiment II were more emotional than those in Experiment I.

The interaction of days by arms, although also a participant in a higher-order interaction, is of interest for comparison to the results obtained in Experiment I. The interaction of days by arms for Experiment I is presented at the left of Figure 2; the comparable data for Experiment II are presented at the right of Figure 2. As is apparent from examining these two plots, the clear differences seen among the arms in terms of frequency of entrance during days 1–4 of testing in Experiment I disappeared with continued testing. The same set of differences seen on days 1–3 of testing in Experiment II were maintained for a much longer time. The order of frequency of entering the arms of the maze remains: the black arm was

entered most often, the white one avoided, and the striped and checkered arms entered with intermediate frequencies.

The effect of days by periods by arms is illustrated in Figure 4, which shows the data on frequency averaged over days 1–5 and days 6–10. The trends seen over all 10 days were strong and continuous, but by combining data the essential elements of the interaction can be simplified and more clearly depicted, despite the fact that magnitudes are understated. The tendency for most of the arms to be entered during the first period on test days 1–5 was accentuated on days 6–10. The tendency for more equality among arms in terms of frequency of entrance with continued testing can also be seen in this figure.

Examination of the data for the interaction of preplacement by periods by arms revealed that the effects produced by differential preplacement were small. Relative to subjects preplaced in the black box, there was a tendency for subjects preplaced in the gray box to enter fewer arms, the lower rates occurring to different arms in the three periods.

Analysis of the data on location using the same factors revealed a significant main effect of arms and interactions of days by arms [$F(27, 1{,}134) = 19.53$] and periods by arms [$F(6, 252) = 3.35$].

The interaction of days by arms, shown at the right of Figure 3, is comparable to the analogous data for Experiment I, shown at the left of Figure 3. In both experiments, there was a decreasing likelihood of an animal being found in the black arm and an increasing likelihood of its being found in another arm. In comparison to Experiment I, however, the preference for the black arm and the avoidance of the white one were stronger and were slower to change over days. The data on location are thus consistent with the data on frequency of entrance in showing the animals in Experiment II to be more emotional and limited in exploratory behavior than the animals in Experiment I.

The interaction of periods by arms of the maze is shown in Table 1. As can be seen in this table, there was an increasing likelihood of an animal being found in the black arm and a decreasing likelihood of it being found in any other arm over the three successive 1-min periods. The same interaction was seen in Experiment I, but it was less extreme. Again, the data fit an interpretation that the animals in Experiment II were more emotional than the animals in Experiment I.

Shock (day 11)

Analysis of the data on *frequency* with factors of preplacement, shock, periods, and arms revealed a significant effect of shock [$F(1, 40) = 12.94$] and an interaction of shock by periods [$F(2, 80) = 5.64$]. These two sig-

nificant effects indicate not only that the shocked animals were less active than the nonshocked animals but also that the effect of shock was greatest in the first period of testing and weakest in the third.

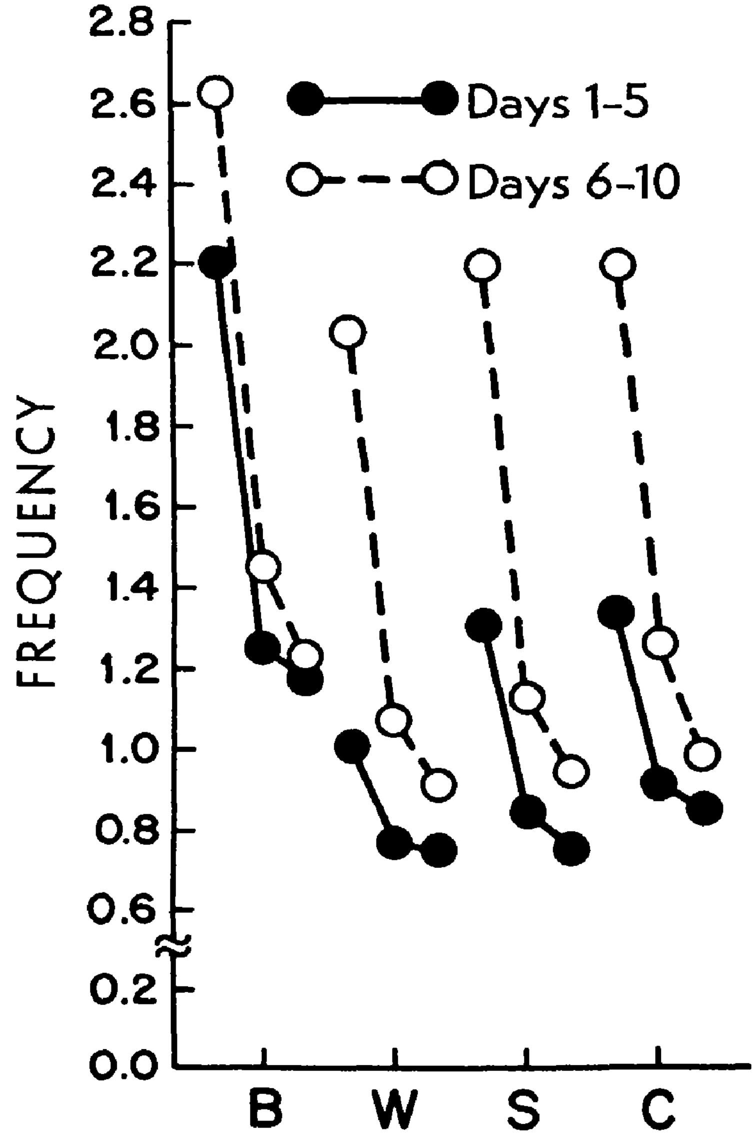


Figure 4. Frequency during each 1-min period for testing days 1–5 and 6–10 (the interaction of days by periods by arms) with preplacements combined: Experiment II

There was also an effect of arms [$F(3, 120) = 16.97$] and an interaction of periods by arms [$F(6, 240) = 2.91$]. This interaction provides the same information as that from the parallel interactions for data over days 1–10 of the present study and days 1–12 in Experiment I.

A similar analysis of the data on *location* revealed only an effect of arms [$F(3, 120) = 30.04$], which is similar to the effect observed in the data for days 1–10.

In general, these analyses indicate that the major effect of shock was on activity rather than preference for a given arm of the maze. In addition, the *place* where the shock was given had no appreciable effect on the measures of either activity or location.

Recovery (days 12–16)

The data on *frequency* were analyzed with factors of preplacement, shock, days, periods, and arms of the maze and revealed main effects of periods and arms and two-factor interactions of shock by days, periods by days, and periods by arms. The interaction of all five factors was also significant [$F(24, 960) = 1.72$]. Despite the presence of this higher-order interaction, the interaction of shock by days [$F(4, 160) = 4.62$] deserves special attention. The nonshocked animals continued to show the decreasing trend in activity that had begun about day 7 (see Figure 1). The shocked animals, however, showed a marked decrease in activity on day 12, with a gradual recovery of activity on subsequent days. Given that activity may be indicative of emotionality, perhaps two processes are functional: during initial exposure to the maze, there is a gradual increasing activity as the emotional response is habituated; at the same time, however, the increased activity brings the animal into greater contact with the elements of the maze, providing the opportunity for satiation to these elements and thereby decreasing exploration and activity. Shock would decrease the activity of the animal below this satiation point.

To help clarify the five-way interaction, the data for the shocked and nonshocked animals were analyzed separately with factors of preplacement, days, periods, and arms. The analysis for the *nonshocked* animals revealed main effects of periods and arms and an interaction of periods by arms [$F(6, 120) = 2.35$]. These effects, except for being at a lower overall level, are virtually identical in topography to those for days 6–10 (Figure 4). There was also a significant main effect of days [$F(4, 80) = 3.78$], confirming the decline in activity over the final days of testing. There was, in addition, a significant interaction of preplacement by days by periods [$F(8, 160) = 2.12$]. Animals preplaced in the gray box showed more activity in the first period than those preplaced in the black box. In

addition, the latter animals showed a relatively constant amount of activity in the first two periods, with some decrease in the third over days. In contrast, the rats preplaced in the gray box showed a decrease in activity over days in the first period and weaker decreasing trends for the other two periods. It would appear that the rats preplaced in the black box reached a stable level of activity by day 12 of testing, whereas the slightly more active rats preplaced in the gray box did not as soon reach stability in their topography of activity.

The parallel analysis of the data for the *shocked* animals yielded significant main effects of periods and arms and a significant interaction of preplacement by days by periods by arms [$F(24, 480) = 1.60$]. This interaction is presented in Figure 5. It should be noted that since the data for days 12 and 13, and for days 15 and 16, have been averaged, this figure is a conservative representation, not only of the magnitude of suppression of activity, but also of the extent of recovery. Several points should be noted. First, the overall levels of activity on the first two days after shock (days 12 and 13, Figure 5) were quite suppressed compared to those be-

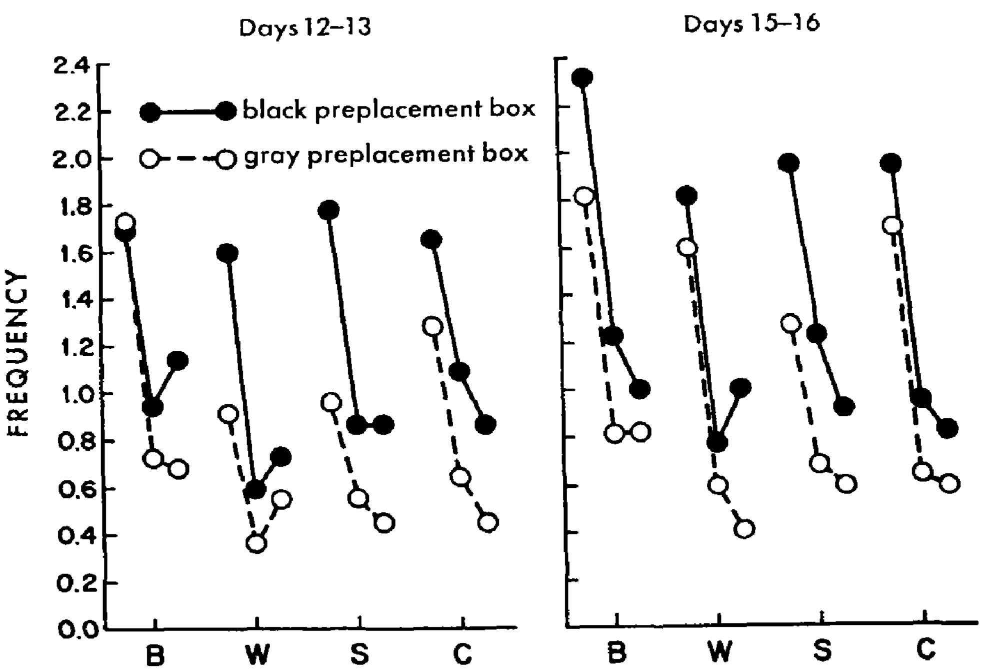


Figure 5. Frequency during each 1-min period for days 12–13 and 15–16 by rats shocked in the black or gray preplacement box on day 11 (the interaction of preplacement by days by periods by arms): Experiment II

fore shock (days 1–10, Figure 4). Second, animals preplaced in the gray box showed less activity than animals preplaced in the black one, in contrast to the trend shown by the nonshocked animals, and the rats preplaced in the gray box showed less increase in activity than those preplaced in the black box. Note also the equal frequency of entrance into all arms on days 12 and 13 by rats preplaced in the black box, indicating decreased entrance into the 'preferred' arm of the maze, and the strong recovery of frequency of entrance into this 'preferred' arm by days 15 and 16. By contrast, the rats preplaced in the gray box showed suppression and recovery of frequency of entrance into the white arm and a suppression of frequency of entrance into the striped arm. Perhaps these results indicate a generalization of fear from the simple gray preplacement — and for these rats, shocking — box to the white arm and to the simpler arm of intermediate intensity.

The data on *location* every 10 sec were also subjected to analysis of variance with factors of preplacement, shock, days, periods, and arms of the maze. This analysis yielded a significant main effect of arms and significant interactions of days by arms and periods by arms. In addition, there was an interaction among all five factors [$F(24, 960) = 1.54$]. To clarify this interaction, data for shocked and nonshocked animals were analyzed separately. Analysis for the *nonshocked* animals resulted in a significant effect of arms [$F(3, 60) = 16.98$] and an interaction of days by periods by arms [$F(24, 480) = 1.94$]. Examination of the data revealed that this interaction represents a continuation of the pattern of results seen at the right of Figure 3.

Analysis of the data on location for the *shocked* animals revealed a significant main effect of arms and interactions of days by arms, periods by arms, and preplacement by days by periods by arms [$F(24, 480) = 1.97$]. Trends representative of this interaction are presented in Table 3. In general, these trends substantiate those seen in the data on frequency. Animals shocked in the black box initially showed an avoidance of the black arm and an increased preference for the striped one. The rats preplaced in the black box showed a rapid recovery from the effects of shock, regaining their preference for the black arm and showing equal preferences for the others. Those animals shocked in the gray box showed a more prolonged and pronounced disruption of their preference hierarchy. They initially showed an enhanced likelihood of being found in the black and checkered arms of the maze and a reduced likelihood of being found in the white and striped arms. These trends showed very slow recovery over the final days of testing.

Table 3. Location during each 1-min period of testing days 12–13 and 15–16, for rats shocked in the black or gray preplacement box on day 11 (the interaction of preplacement by days by periods by shock); Experiment II

	Arm of maze			
	Black	White	Striped	Checkered
Days 12–13				
Black preplacement box				
First period	1.68	1.14	1.91	1.27
Second period	2.27	.45	2.00	1.27
Third period	1.96	1.14	1.86	1.05
Gray preplacement box				
First period	2.59	.64	.82	1.95
Second period	2.77	.55	.91	1.77
Third period	3.23	.77	.82	1.18
Days 15–16				
Black preplacement box				
First period	2.55	1.05	1.00	1.41
Second period	2.32	1.14	1.45	1.09
Third period	2 68	1.14	1.23	.95
Gray preplacement box				
First period	2.45	1.32	.82	1.41
Second period	3.32	.64	.82	1.23
Third period	3.14	.77	.91	1.18

DISCUSSION

In general, the overall results suggest that if rats are given daily exposure for three consecutive 1-min periods over 10 or more days to a maze whose arms differ in brightness and complexity, certain changes will occur in the animals' activity and in their pattern of exploration of the maze. Initially, the animals will be relatively inactive and this low level of activity will be equally distributed throughout the time the animals are in the maze. The animals will enter the black arm most frequently and the white arm least frequently, and congruently, will most likely be found in the black arm and least likely in the white one at any time during the testing period.

As testing continues over days, there will be an increase in activity, this activity becoming concentrated in the first 1-min period of exposure to the maze on each daily trial. Although the animals will continue to show a preference for the black arm and an avoidance of the white one, there will be an increased frequency of entrance into the other arms and an

increased likelihood of being found in them, at least during the first 1-min period.

Habituating the animals to handling, as in Experiment I, will accelerate these changes in activity and exploratory behavior. If, as Thompson and Lippman (1972, 1975) have suggested, habituation to handling before testing in a maze reduces overall emotionality and fearfulness, then the changes described above should occur more rapidly in habituated animals than nonhabituated animals. The results of Experiments I and II support this suggestion. As already noted, the differences in the rats' performance between the two experiments were most obvious during the first two days of testing and were virtually absent by the end of the first week of testing.

In addition to the normal handling concomitant with each daily test trial, the animals in Experiment II were preplaced in either a black or gray box, which also involves handling. Considering that the animals in Experiment II had not received habituation to handling, this additional handling could have increased the animals' emotionality and fear, thus further attenuating activity and exploratory behavior.

Observations from previous experiments (Thompson and Lippman, 1972, 1975) have indicated that the black arm of a maze is approached more and the white one is avoided more than any of the others. The minor effects of preplacement observed in Experiment II can be accounted for if it is assumed that tendencies to approach or avoid are determined by the amounts of fear evoked by the various arms of a maze and that these different effects of fear can generalize. Relative to rats preplaced in the gray box, rats preplaced in the black box showed less extreme differences in frequency of entrance into the various arms of the maze and a slightly higher level of activity, all indicating slightly lower levels of fear.

The gradual concentration of exploratory behavior and activity in the first 1-min period of each daily trial in both experiments is consistent with the hypothesis that the animal explores to reduce fear (Halliday, 1967; Lester, 1967). Since the handling required to put the animal in the maze each day would induce fear (more on earlier trials than on later trials), the animal would then explore to reduce the fear. As trials continue and as the animals become habituated to handling, the amount of fear evoked by the handling would decrease and would decrease at an accelerated rate. Thus, the animals would engage in exploring only during the initial moments of contact with the maze each day.

Shocking the animals before putting them in the maze reinstates fear and the animals' behavior topography is similar to that shown upon initial exposure to the maze: activity is low and there is a strong preference for

the black arm and avoidance of the others, especially the white one. The extent to which the animals revert to previous patterns of behavior and the time it takes for them to recover from the effects of the shock-induced fear probably depend on the intensity and number of shocks received, as well as subsequent opportunity to explore the maze. In Experiment I, the rats showed a clear-cut return to the behavior shown on the initial test days and little recovery from the shock-produced fear on the first two days after shock. The effects for Experiment II, in which the rats received fewer shocks, were less clear-cut and some recovery was apparent over the four days after shock. The results of Experiment II were complicated by shocking the rats in a box either similar or dissimilar to any of the arms of the maze.

Animals shocked in the gray preplacement box showed a return to the behavior they displayed at the beginning of the experiment. In addition, there appeared to be some generalization from the gray box to the white arm of the maze, the rats shocked in the gray box showing avoidance of the white arm on the first two days after shock, with some recovery on the last two days. The activity of the animals shocked in the gray box was suppressed more than that of the animals shocked in the black box. It is plausible to assume that the rats preplaced in the gray box had retained more fear over the previous days of testing and that this fear summated with the fear induced by the subsequent shock in the gray box.

The animals shocked in the black preplacement box did not show an immediate return to the pattern shown on the first day of testing. Overall activity was reduced, but generalization of the shock-induced fear, which had been paired with the black preplacement box, to the black arm of the maze initially reduced entrances into that arm. However, by the last two test days there had been a rapid recovery from the fear elicited by the shock and the behavior of these animals was almost identical to their behavior just before they were shocked.

The data from these two experiments seem compatible with the hypotheses of Halliday (1967) and Lester (1967). Factors that reduce fear will at first increase and later decrease activity and exploratory behavior. Brief reinstatement of fear will at first decrease activity and exploratory behavior, such that the animal will behave much as it did when first confronted with those particular stimuli.

Notes

The authors may be reached at the Department of Psychology, Western Washington University, Bellingham, Washington 98225. Received for publication June 20, 1976; revision, October 22, 1976.

References

Aitken, P. P., and Sheldon, M. H. 1970. Electric shock and rats' preference for the familiar arm of a maze. *British Journal of Psychology* 61:95–97.

Berlyne, D. E. 1960. *Conflict, arousal, and curiosity.* New York: McGraw-Hill.

Bronstein, P. M. 1972. Open-field behavior of the rat as a function of age: Cross-sectional and longitudinal investigations. *Journal of Comparative and Physiological Psychology* 80:335–341.

DeNelsky, G. Y., and Denenberg, V. H. 1967. Infantile stimulation and adult exploratory behavior in the rat: Effects of handling upon visual variation seeking. *Animal Behaviour* 15:568–573.

Fiske, D. W., and Maddi, S. R. 1961. A conceptual framework. In *Functions of varied experience,* ed. D. W. Fiske and S. R. Maddi. Homewood, Ill.: Dorsey Press.

Halliday, M. S. 1967. Exploratory behavior in elevated and enclosed mazes. *Quarterly Journal of Experimental Psychology* 19:264–271.

Haywood, H. C., and Wachs, T. D. 1967. Effects of arousing stimulation upon novelty preference in rats. *British Journal of Psychology* 58:77–84.

Lester, D. 1967. Sex differences in exploration: Toward a theory of exploration. *Psychological Record* 17:55–62.

Lippman, L. G., Galosy, R. A., and Thompson, R. W. 1970. Passive-avoidance learning in gerbils and rats. *Journal of Comparative and Physiological Psychology* 73:269–273.

Montgomery, K. C. 1955. The relationship between fear induced by novel stimulation and exploratory behavior. *Journal of Comparative and Physiological Psychology* 48:254–260.

Russell, P. A. 1973. Relationships between exploratory behavior and fear: A review. *British Journal of Psychology* 64:417–433.

Sheldon, M. H. 1968. The effect of electric shock on rats' choice between familiar and unfamiliar maze arms: A replication. *Quarterly Journal of Experimental Psychology* 20:400–404.

Thompson, R. W., and Lippman, L. G. 1972. Exploration and activity in the gerbil and rat. *Journal of Comparative and Physiological Psychology* 80:439–448.

Thompson, R. W., and Lippman, L. G. 1975. The effect of age and postweaning and adult handling habituation on activity and exploration in the rat. *Bulletin of the Psychonomic Society* 5:285–288.

AMERICAN JOURNAL OF PSYCHOLOGY
September 1977, Vol. 90, No. 3, pp. 383–406

The role of opportunities for recall in learning to retrieve

Henry M. Halff
University of Illinois at Urbana-Champaign

Three theories of learning hold that test trials are equally effective on errors or successes; that test trials are effective only for successfully recalled material; or, by Restle's strategy-selection model, that presentations are effective only following failures of recall. The implications of these three theories were developed for the parameters of a two-stage Markov model of learning. An application of the model to the RTT paradigm in free recall provided support for the strategy-selection theory, as did two experiments involving error-contingent presentation. The paper concludes with a theory relating organizational aspects of free recall to the strategy-selection theory.

This paper is addressed to the process of learning to retrieve learned verbal material. Retrieval itself may be viewed as involving a plan or strategy for recovering the appropriate material in the particular situation for recall. Most theorists suggest that these plans operate in one of two ways. Theorists such as Anderson (1972), Kintsch (1970, 1972, 1974), and Shiffrin (1970) argue that retrieval is the exercise of a fairly general plan whose basic nature is essentially independent of the particular stored material. But other types of evidence (Bower, 1970; Thompson and Tulving, 1970; Tulving and Osler, 1968; Tulving and Thompson, 1973) indicate that retrieval may involve more specialized procedures whereby people often develop plans for retrieval that take advantage of the specific configuration of stored material. In fact, both general and specific plans play a role in recall. Specific plans usually take some time to develop, so that subjects must be using some sort of general technique to recall newly learned material. Incidental learning is further evidence of people's ability to retrieve in the absence of specific plans. On the other hand, it is hardly likely that subjects in a typical verbal-learning experiment will never de-

velop a plan for retrieving the material they are required to recall. Indeed, such specific plans will probably develop whenever a person is required to memorize specific material and is given adequate time and practice.

Learning to retrieve, the process of developing specific plans for retrieval, is the focus of the following studies, but before any questions may be asked, we need a general model of the learning process, one which adequately describes the data and within which the process of learning to retrieve can be examined. Some concern with this problem already exists in the literature, and a promising descriptive model of the learning process has been developed by Greeno (1970), Greeno, James, and DaPolito (1971), and Humphreys and Greeno (1970). The studies presented here were designed within the framework of this general model, in hopes of answering some important questions about learning to retrieve.

My interpretation of Greeno's theory is that learning to retrieve involves the transition from the use of a general heuristic to the use of a specialized algorithm. Heuristic retrieval is essentially a set of procedures that maximizes the chances of successful recall within the constraints of the situation. A detailed description of such procedures is inappropriate at this point in the paper, but assume that one defining characteristic of heuristic retrieval is its fallibility. That is, stored material will not always be retrieved if heuristic retrieval is being used. By contrast, people eventually come to use a specialized algorithm that always retrieves the target material. The exact nature of such algorithms is an important issue, but, as with heuristic retrieval, its discussion is more appropriate later on. For the moment, assume that one defining characteristic of algorithmic retrieval is its infallibility. That is, within the framework of the experiments considered here, algorithmic retrieval, once developed, will always retrieve the target responses.

The actual learning of a list of items over a series of trials can be characterized by a two-stage model. In the first stage, each item is stored in memory after some number of trials; in the second stage, further trials are spent developing a plan for retrieving the stored material. Heuristic retrieval is used between storage and development of an algorithm, and algorithmic retrieval is used thereafter. To be specific, each item is in one of three states on any particular trial. Items are in state O prior to storage. Stored items are in state H prior to development of an algorithm and in state L thereafter. State H may be broken down into two substates, E and S, depending on whether heuristic retrieval failed or succeeded, respectively, on that trial. Recall is, of course, impossible in state O and certain in state L.

In many situations, transitions between the states defined above may

be described by the following finite Markov process, which is also illustrated in Figure 1.

$$P = \begin{array}{c} L(n) \\ E(n) \\ S(n) \\ O(n) \end{array} \overset{\begin{array}{cccc} L(n+1) & E(n+1) & S(n+1) & O(n+1) \end{array}}{\begin{bmatrix} 1 & 0 & 0 & 0 \\ d & (1-d)q & (1-d)p & 0 \\ c & (1-c)q & (1-c)p & 0 \\ ab & a(1-b)q & a(1-b)p & 1-a \end{bmatrix}},$$

$$P[L(1)\ E(1)\ S(1)\ O(1)] = [ab\ a(1-b)q\ a(1-b)p\ 1-a],$$

[Equation 1]

where $p = 1 - q$. Note that on each trial an unstored item is stored with probability a, thence passing directly to state L with probability b or remaining in state H with probability $1 - b$. Items in state H pass to state L with probability c on each trial that heuristic retrieval succeeds and with probability d on each trial that heuristic retrieval fails. Finally, assume that heuristic retrieval has a constant failure rate, q. Detailed mathematical discussions of this model may be found in Greeno (1968) and Halff (1976).

Two questions are appropriate at this juncture. First, does the model adequately describe the learning process, and second, how can the model best be exploited to discover the nature of the underlying processes? Turning to the first question, the model has been applied to paired-associates learning by Humphreys and Greeno (1970), Greeno (1970), Greeno et al. (1971), and Pagel (1973), and to free recall by Heine (1970), Kintsch and Morris (1965), and Waugh and Smith (1962). In all cases, the goodness of fit of the model was acceptable, and the first-stage and sec-

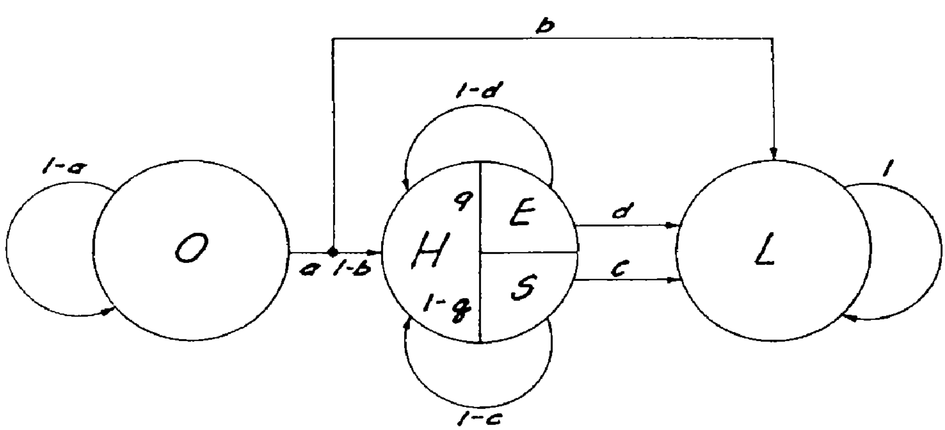


Figure 1. Graphic representation of the two-stage Markov model of learning

ond-stage transition parameters reacted appropriately to manipulations designed to affect storage and learning to retrieve respectively.

In order to see how the two-stage model can help discover how people learn to retrieve, we may turn to another empirical observation related to learning to retrieve, that is, the effect of test trials or opportunities for free recall. Several investigators have found that test trials are equally if not more beneficial to learning than presentations. Lachman and Laughery (1968) and Tulving (1967) were the first investigators to examine this phenomenon, and since then, most studies of this effect have presented evidence indicating that test trials increase the retrievability of the material (Birnbaum and Eichner, 1971; Donaldson, 1971; Hogan and Kintsch 1971; Rosner, 1970). There are at least three important classes of theories of learning to retrieve. Each of these has important implications for the two-stage model, and each is relevant to the beneficial effects of test trials.

The simplest of these is a *general-potentiation theory*,[1] holding that each test trial has an overall potentiating effect regardless of performance on that trial. This explanation, although not a priori attractive, seems to be the implicit assumption of investigators, such as Lachman and Laughery (1968) and Tulving (1967), who are interested in global comparisons of the effects of test trials with those of presentations. Such global comparisons would be meaningless if the effects of test trials were dependent on performance on the test. There are several possible justifications for a general-potentiation theory. First, the effects of tests might be purely motivational. The more tests a subject anticipates, the harder he might study. Second, test trials might have their effects on the consolidation of a single structure representing the list in memory, and this consolidation might take place independent of the particular recall protocol the structure produces on the test. Finally, Izawa (1971), in a theory of paired-associates learning, suggests that test trials have an overall potentiating effect on the subsequent presentation. If these potentiating effects and any other effects of test trials are independent of performance on the test, then the parameters of the two-stage theory illustrated in Figure 1 should be such that $b = c = d$; for present purposes, this requirement will be the defining characteristic of the general-potentiation theory.

A more plausible conception of the role of test trials would limit some or all of their beneficial effects to items successfully recalled on the test. There are at least two theories in the literature that make such a suggestion. Izawa (1971) proposes, in addition to a general potentiation, a beneficial effect of test trials on the retention of items successfully recalled, this latter because effective retrieval cues for these items are maintained in a state of high availability through the retrieval process. Adams and

Bray (1970, p. 392) make a similar suggestion in their closed-loop theory of paired-associates learning. In their own words, "it is repetition of the response through rehearsal, either overt or covert, which is the basic operation for strengthening of a verbal response." Bjork (1975) makes a similar suggestion, but he proposes that only particular types of retrieval operations are effective in learning. To my knowledge, the earliest appearance of this type of principle in the psychological literature was Thorndike's *law of exercise* (cited in Hilgard and Bower, 1975, p. 33). We will use Thorndike's terminology and refer to any theory holding that test trials have special beneficial effects for material successfully retrieved on those trials as subscribing to the law of exercise. (It is worth mentioning that such theories do not exclude learning on presentations after errors or even a general-potentiation factor associated with test trials.) In terms of the parameters of the two-stage model, reference to Figure 1 suggests that the law of exercise requires that $b = d$ and that $c > d$.

Finally, there is a theory holding that failure to recall is the crucial event on a test trial. This theory was originally proposed for cue learning by Restle (1962), and it has been extended to some paired-associates situations by Greeno and Scandura (1966), Polson, Restle, and Polson (1965), Polson (1972), and Restle (1964, 1965). This *strategy-selection theory* holds that the subject uses one of a number of recall strategies or mnenomic codes on each trial. A version of the theory proposed by Restle (1965) is that any such strategy may be either fallible, failing on some random proportion of the trials, or infallible. The subject randomly selects a strategy for each item when it is stored (i.e., when it enters state H), and each strategy is maintained until it fails. Thus, on each presentation the subject determines whether or not he recalled the item on the previous trial, and if he did not, selects another strategy. This theory can be related to the two-stage model by assuming that the number of strategies is large or that strategies are sampled without replacement, that the proportion of infallible strategies is d, and that the probability of successful retrieval with a fallible strategy is p. We may then identify state H with the use of a fallible strategy and the passage from state H to state L with the sampling of an infallible strategy. This transition will then occur with probability d after failures of retrieval and never after successful retrieval. Hence, Equation 1 represents the process with $b = d$ and $c = 0$.

To determine which if any of these three theories is tenable is clearly a crucial step in understanding the nature of learning to retrieve, and this determination, for free recall, is the purpose of this paper. Since we have already established a specific relationship between each theory and the parameters of the two-stage model, a natural next step would be to obtain

estimates of the parameters. Standard inference techniques could then be used to test the implications of the theories as statistical hypotheses about the parameters. Unfortunately, the parameter of interest, namely the ratio of c to d, cannot be estimated from the data produced by the process specified in Equation 1. The reason for this, given in greater detail in Greeno (1968), is that the likelihood of such data is solely a function of the three parameters a, $(1 - c)p$, and $(1 - d)q$. Hence, it is possible to obtain unique best estimates only of these three parameters or functions of them. Unfortunately, the ratio, c/d, is not a function of the above three parameters and hence cannot be estimated.

To solve this problem, we can extend the two-stage model to the RTT paradigm in which *two* tests are given after each presentation. This extension provides data enough to yield estimates of all of the crucial parameters and hence allow an evaluation of the three theories under consideration. But to accomplish this extension to the RTT case, we require a more detailed account of learning to retrieve.

Let us develop such an account. First, note that the three theories, taken together, suggest that learning to retrieve may occur at any of three different points in a trial, namely, when an item is successfully retrieved, on a presentation following successful recall, and on a presentation following a recall failure. According to the law of exercise, successful retrieval itself results in learning to retrieve for some proportion, say c', of the items in state H. Both the law of exercise and the general-potentiation theory allow learning to retrieve on presentations following successful recall; let d' be the proportion of items so learned. Finally, all three theories allow learning to retrieve with probability d on presentations following failures of recall. This general account yields a slightly more detailed description of the fate of items in state H on any trial. As is illustrated in Figure 2, a proportion p of such items will be retrieved on any test, and c' of those retrieved items will pass to state L on that test. Of the successfully retrieved items that do not pass to state L upon retrieval, d' will pass to state L on the subsequent presentation. Finally, of the items not retrieved on that test, d will pass to state L on the subsequent presentation. From Figures 1 and 2 then, it is clear that the parameter, c, which reflects learning to retrieve after successes, is a function of c', namely, the effect of the retrieval itself, and d', the effect of the subsequent presentation. That is,

$$c = c' + (1 - c')d'.$$

[Equation 2]

According to this more general account, the crucial parameters for

distinguishing between the three theories are c', the probability that recall results in learning to retrieve, and d', the probability that presentations following recall result in learning to retrieve. The general-potentiation model assigns no special role to performance on tests and hence requires that $c' = 0$ and $d' = d$. The law of exercise holds that successful recall itself may result in learning to retrieve but that performance has no effect on the subsequent presentation — that is, that $c' > 0$ and $d' = d$. Finally, the strategy-selection theory holds that the only event which can result in learning to retrieve is a presentation following a failure of recall — that is, that $c' = d' = 0$.

To extend this model from the RT case, illustrated in Figure 2, to the RTT case, similarly illustrated in Figure 3, requires few additional assumptions. The storage process is the same in both paradigms; that is, unstored items are stored with probability a on each presentation. To allow for forgetting, assume that items in state H are retrieved with probability p_1 on the first test and with probability p_2 on the second test but that such retrieval is dependent only on the state of the process at the time of testing. (Let q_1 and q_2 represent the corresponding complementary probabilities.) Concerning learning to retrieve on tests, assume that any successful retrieval results in passage to state L with probability c'. For learning on presentations, note that the parameters d and d' are applied following errors and successes respectively in the RT case, to allow for the strategy-

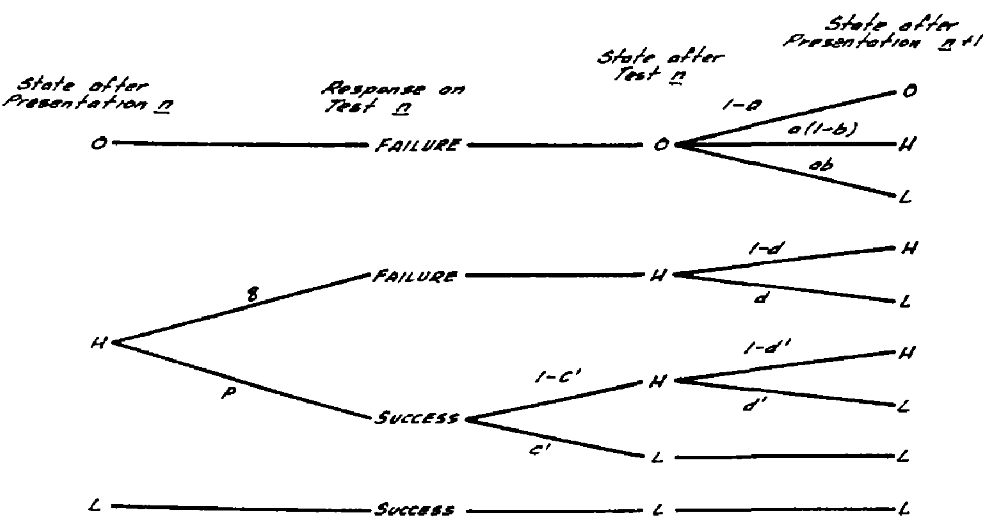


Figure 2. Tree diagram of a trial according to the two-stage model

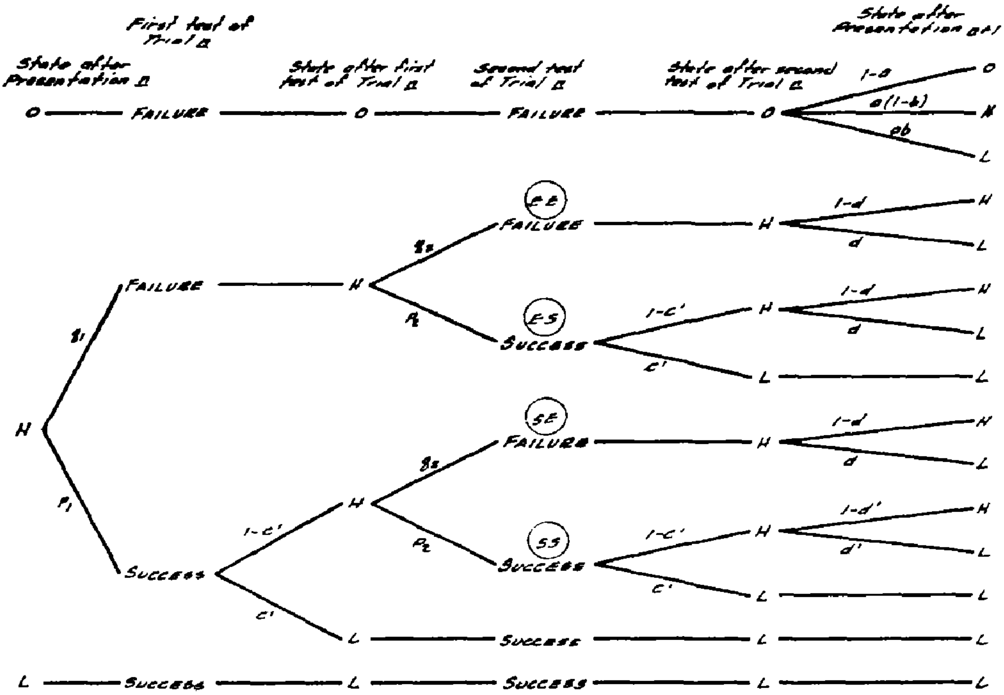


Figure 3. Tree diagram of a trial in the RTT paradigm according to the two-stage model

selection theory's requirement that strategies are selected only after failures of recall. In the RTT case, the strategy-selection theory should require that such reselection will occur on presentations after a failure of recall on *either* test but that strategies will be retained if successes occur on *both* tests. This suggests the assumption that learning to retrieve will occur on a presentation with probability d if either of the previous tests had a failure of recall and with probability d' if both tests resulted in successful recall.

To see how this rather complicated process relates to the question at hand is not as difficult as it may seem. Note that the main issues involve the parameters c' and d'. If learning occurs by exercise, then c' will be greater than zero. If the effects of presentation are independent of performance on the previous trial, then d' and d will be equal, but if learning occurs only after recall failures, then d' will be zero. Now consider the implications of each of these conjectures for the process illustrated in Figure 3 and, in particular, for the probabilities of reaching criterion at

each of the four points labeled *SS*, *SE*, *ES*, and *EE* which indicate the status of each item in state *H* on the second test of a trial. If $c' > 0$, then items at point *ES* have a better chance of reaching criterion than do those at points *SE* or *EE*, but if $c' = 0$, then items at all three of these points should have the same chance of reaching criterion. Similarly, if $d' = d$, then items at point *SS* should have the same chance of reaching criterion as items at point *ES*, but if $c' = d' = 0$, as the strategy-selection theory requires, then it should be impossible to reach criterion directly from point *SS*. With this as introduction, we may turn now to an experiment designed to provide the data necessary to test the conjectures just described.

EXPERIMENT I

METHOD

Subjects

The subjects were 14 men and 14 women. They served as a part of a course requirement in an introductory psychology course at the University of Illinois.

Materials and apparatus

The materials consisted of 12 semantically unrelated four-letter nouns, none of which began with the same letter. These words were chosen from the set used by Heine (1970). The materials were presented to the subjects via slides projected on a screen about 3 m in front of the subjects.

Procedure

Subjects were run in groups of four, but a certain amount of isolation was achieved by placing subjects in booths open only at the front, back, and top.

The subjects received two practice trials on a list of ten three-letter nouns followed by twelve trials on the list described above. Each trial consisted of a presentation of the list in a different random order followed by two consecutive recall cycles. The words were presented one at a time at a 1.5-sec rate. Each recall cycle consisted of a 40-sec period for addition and a 90-sec period for recall. During the period for addition, the subjects computed the sums of as many lists of eight digits as they could in the allotted time. During the period for recall, they wrote each word they remembered on a separate card, placing the cards in a covered box as they were filled in.

RESULTS AND DISCUSSION

Since each subject learned twelve items, there were 336 combinations (28 subjects × 12 items). The data for this study consist of the response protocols for each of these combinations. Each such protocol specifies whether or not the particular item was recalled by the particular subject on each of the two tests of each trial.

A parametric analysis

Our first concern is with parameter estimation; specifically with finding those values of the parameters of the two-stage model — a, c', d, d', q_1, and q_2 — that give the best account of the data. To do this, the method of maximum likelihood was used. First, an expression for the joint likelihood of the 336 protocols in terms of the parameters of the model was derived. In deriving this function, it was assumed that each sequence was an independent draw from the process defined in Figure 3 and that the same set of parameters applied to all 336 protocols. (See Atkinson, Bower, and Crothers, 1965, for further information on this technique.) Next, a computer program, Chandler's (1969) STEPIT, was used to find those values of the parameters that maximized this likelihood function. The resulting values were

$$a = .679,\ c' = 0,\ d' = 0,\ d = .648,\ q_1 = .296,\ q_2 = .353. \qquad \text{[Equation 3]}$$

Note that these estimates conform exactly to the implications of the strategy-selection theory that $c' = 0$ and $d' = 0$. Thus, of the three theories under consideration, the data conform best to the strategy-selection theory. What remains to be seen, then, is whether or not the other two theories, both of which require that $d' = d$, also give a plausible account of the data. In other words, we need a test of the statistical hypothesis that $d' = d$.

A theorem of Wilks (1962, chapter 13) can be used to provide an approximate statistical test of this hypothesis. Wilks's theorem applies to cases in which a null hypothesis is some number, n, of linear restrictions on a more general alternative hypothesis. In this case, the hypothesis that $d' = d$ imposes a single linear restriction on the parameter space of the model, so that $n = 1$. Wilks suggests comparing the maximum likelihood achievable under the restriction(s) to the maximum likelihood achievable in the unrestricted case. If we let L_0 be the former likelihood, L_1 be the latter, and λ be the ratio, L_0/L_1, then according to Wilks's theorem, the limiting distribution of $-2 1n\lambda$ under the null hypothesis is χ^2 with n degrees of freedom. However, if the null hypothesis is false, L_0 will be much smaller than L_1, and $-2 1n\lambda$ will, on the average, be larger than $\chi^2(n)$.

For the data from this experiment, L_1 is simply the likelihood of the data under the parameter estimates given in Equation 3; using these estimates, $-2 1nL_1 = 2071.87$. To obtain L_0, STEPIT was used to find the maximum likelihood under the restriction that $d' = d$; the result was that $-2 1nL_0 = 2099.71$. The difference between these two figures is, of course,

highly significant [$\chi^2(1) = 27.84, p < .001$]. Thus, not only does the strategy-selection theory appear to be the most attractive of the three theories, but the two alternatives to it can be rejected on statistical grounds.

Any conclusions drawn from these results must, however, be considered in the light of the goodness-of-fit of the model, and unfortunately, there are clear deviations from good fit in many aspects of the data. To summarize, the model could not account for the small variance of the initial error run, but the requirement of the model that the number of initial errors be independent of subsequent events did seem to be satisfied. With respect to precriterion responses, the model could not account for a lack of stationarity and independence of responses between the first success and last error. Nor could it account for the disproportionately low error rate on the trial of the first success. Finally, two requirements concerning the location of the trial of the last error were satisfied in the data. Specifically, the probability that any error trial was the last was independent of the number of previous trials and errors, and the probability of reaching criterion at the point labeled *SE* in Figure 3 was the same as the corresponding probability for point *EE*.

Some nonparametric analyses

These discrepancies in fit indicate that inferences based on the estimates in Equation 3 should be interpreted with caution. Fortunately, however, there are certain statistics that permit a nonparametric evaluation of the important theoretical issues and thus at least partially avoid the problems of poor fit. Recall that the two crucial parameters of the model are c', which reflects learning due to successful recall itself, and d', which reflects learning on presentations following successful recall. Turning first to the former parameter, we can see from Figure 3 that c' will be nonzero if and only if the probability of reaching criterion after occurrences of state *ES* exceeds the corresponding probability for states *EE* and *SE*. The relative frequencies corresponding to these events are .712 ($N = 59$) and .592 ($N = 108$) respectively. While there is some indication that c' exceeds zero, the difference between the two probabilities is not significant [$\chi^2(1) = 1.85$].[2] It is worth noting that these statistics are based on aspects of the data that were well fit by the model.

Inferences about d' are not so easily arrived at or interpreted. Recall from the introduction that information about d' primarily rests on the probability of reaching criterion from point *SS*, or alternatively, on its complement, the probability, f, of at least one error after that point. Unfortunately, those arrivals at point *SS* that do lead to criterion are never

observable. We know that such occurrences must be after the last error, but the actual number of times the process goes through point *SS* after the last error cannot be determined. Nonetheless, f may be estimated indirectly by comparing two ratios. The first, x, is the ratio of the probability of arrival at point *EE* to that of arrival at point *SE*. The second, y, is the ratio of probability of arrival at point *SE* to that of a *precriterion* arrival at point *SS*. From Figure 3, it is apparent that $x = q_2/p_2$, but since only precriterion arrivals at *SS* figure in the denominator of y, $y = q_2/(p_2 f)$. Solving for f, we find $f = q_2/(p_2 y) = x/y$. In these data, there were 29 arrivals at point *EE* and 50 arrivals at point *ES*; the ratio, .580, is an estimate of x. Similarly, there were 58 arrivals at point *SE* and 90 precriterion arrivals at point *SS*, yielding an estimate of .644 for y. The resulting estimate of f is simply .580/.644, or .900.[3]

Now consider the implications of each theory for f, the probability that point *SS* is followed by at least one error. If the strategy-selection theory is correct, then second-stage learning occurs only on errors and arrivals at point *SS* must always be followed by at least one error. This would imply that $f = 1$, and in fact, the estimated value of f is quite close to and not significantly different from 1 [$\chi^2(1) = .05$].[4] If, on the other hand, the law of exercise or the general-potentiation theory applies, then presentations following successes are equally effective as those following errors and $d' = d$. Reference to Figure 3 reveals that this requirement makes states *SS* and *ES* essentially equivalent and thus implies that f is equal to the probability that at least one error will follow an occurrence of state *ES*. This latter probability is estimated by its corresponding relative frequency at .288 ($N = 59$). Using a minimum χ^2 to compare this probability to f results in a significant difference between the two [$\chi^2(2) = 6.93$, $p < .05$]. In interpreting these figures, a warning is again in order concerning goodness of fit. In particular, the frequencies used to estimate and compare x and y do not seem to fit the model or to meet the assumptions required for the use of Pearson's χ^2. Nonetheless, if the marginal probabilities are at all meaningful, they indicate that the strategy-selection theory is in a much stronger position than either of the other theories, which require learning on presentations following successes.

In summary, the initial parametric analysis, which provided the estimates given in Equation 3, indicates that the strategy-selection theory is the strongest theory of the three we considered. This conclusion, however, is tempered by the model's failure to account for the distribution of initial errors and the characteristics of responses between the first success and the last error. Nonparametric analyses of certain aspects of the data yielded results that were similar to the parametric results. In particular,

there was a nonsignificant indication that learning to retrieve may occur as the result of successful retrieval and a strong indication that learning to retrieve does not occur on presentations following successful recall. But these nonparametric analyses only partially avoid the problems of bad fit, and the following experiments were carried out to avoid the difficulties in fit encountered in Experiment I as well as to provide relevant information from another paradigm.

EXPERIMENTS II AND III

Both of these experiments involved the comparison of an error-contingent presentation scheme with the more conventional procedure of presenting items on every trial. This comparison was the sole purpose of Experiment II, and Experiment III had the additional aim of studying the role of short-term memory for trial n responses during presentation on trial $n + 1$.

On the issue of error-contingent presentation, note again that according to the strategy-selection theory, failures of recall cause the subject to select new strategies on the next presentation, whereas successes result in the subject's ignoring the subsequent presentation. This theory therefore holds that the only psychologically effective presentations are those that follow errors; presentations following successes play no role in learning. A strong test of this theory would therefore involve denying the subjects a presentation after successful recall. Items in one condition, condition C, would be presented on every trial, while items in another condition, condition R, would only be presented on the first trial and on trials following errors.

A certain amount of work using interactive presentation procedures is already in the literature (Buschke, 1973, 1974, 1975; Murdock, Anderson, and Ho, 1974; Murdock, Penney, and Aamiry, 1970). Of these studies, Buschke's are of particular interest for his extensive investigation of a condition R in what he calls selective reminding. While he has not done a detailed analysis within the framework of Equation 1, his theoretical analysis is very close to the one used here, and those aspects of the data he does present conform well to the two-stage model. Buschke's work, together with other applications of the model to free recall (Heine, note 1; Kintsch and Morris, 1965; Waugh and Smith, 1962), suggests that a comparison of the two paradigms in terms of the model might be useful.

As was mentioned above, Experiment III had the additional purpose of dealing with the role of immediate memory for responses on subsequent presentation of the material. It is possible that the subject faces each pre-

sentation with some or all of the words from the previous period for recall still in short-term memory; he may thus react differently to presentation of those words than to the presentation of words not in short-term memory. One simple way of examining the role of short-term memory for responses is to place some sort of intervening activity between each period for recall and the subsequent presentation of the list. This manipulation should effectively degrade the subject's short-term memory for his own responses.

METHOD

Subjects

There 44 subjects in Experiment II, and 60 in Experiment III. In both cases, the subjects were taken from the same population as that used in Experiment I.

Materials and design

The materials for both experiments consisted of the same list of nine words, the words obtained in the same manner as those of Experiment I. This list was randomly divided into three three-word sublists, A, B, and X. For half of the subjects, sublist A was in condition R, and sublist B was in condition C; this assignment was reversed for the remaining subjects. In addition, half of the subjects in each group of Experiment III were given 40 sec of addition after each period for recall; the remaining subjects of Experiment III and all subjects of Experiment II proceeded directly from recall to the subsequent presentation.

Procedure

The subjects were run singly or in pairs with each subject serving in a semi-isolated booth. Stimuli were presented and responses were collected on an ADDS Consul 880 CRT/keyboard console under control of an IBM 1800 computer.

Subjects were first instructed on the use of the console and the nature of the task in conjunction with a two-trial demonstration of the procedure. They were told that they were to learn a list of words but that only part of the list would be presented on each trial. They were not informed of any contingency of response and subsequent presentation. Particular care was taken to ensure that the subjects knew that they were to recall all words seen on any previous trial of the experiment. They were also told that the experiment would last until "the computer decided that they had learned the list well enough." Before running the experiment, the subjects were run on a practice list of six three-letter nouns. The procedure for running the practice list was identical to that described below for the real list except that practice was terminated if the subjects had not met criterion after five trials.

The first trial of the experiment started with a randomly ordered presentation of the six words of sublists A and B at a 2-sec rate. Following this presentation, the subjects were required to add pairs of two-digit numbers for 40 sec. This was followed by a 90-sec period for recall, during which the subject entered on the keyboard all of the words he could remember. Each word was erased from the screen as it was entered. The period for recall was followed by another 40-

sec period of addition for half of the subjects in Experiment III. Subsequent trials of the experiment were identical to the first except that any item in condition R that was successfully recalled on some trial was replaced in the presentation protocol of the next trial with a randomly chosen item from sublist X. Thus, items in condition R were presented only on the first trial and on trials following failures of recall. Each subject was run for ten trials or until he had recalled all words from sublists A and B on two consecutive trials.

RESULTS AND DISCUSSION

One subject from each experiment failed to reach criterion within ten trials and their data were excluded from further analysis. As a preliminary analysis of the effects of the manipulations, the mean number of errors per item, the mean trial of the last error, and the mean number of precriterion presentations were computed for each condition; these statistics are presented in Table 1. An analysis of variance on these means revealed that items in condition R required, on the average, more trials but fewer presentations to reach criterion than those in condition C [$F(1, 96) = 8.59, p < .01$ for trials; $F(1, 96) = 11.40, p < .01$ for presentations]. In addition, the effect of sublist, A or B, was significant for the number of errors [$F(1, 96) = 12.11, p < .001$], trial of the last error [$F(1, 96) = 7.07, p < .01$], and number of precriterion presentations [$F(1, 96) = 7.64, p < .01$]. No other effects were significant.

To determine the exact nature of the differences between conditions, data from all six groups of both experiments were combined to form a total of 306 response protocols (102 subjects × 3 items) in each condition. Analyses of the fit of Equation 1 to these data revealed that, unlike the outcome of Experiment I, the fit of the model was good in almost all respects. The only interesting deviation was a failure to meet the assumption that $b = d$ in condition R [using Wilks's λ, $\chi^2(1) = 30.62, p < .001$].

Table 1. Mean numbers of precriterion errors, trials, and presentations per item; conditions R and C, Experiments II and III

	Errors		Trials		Presentations	
	R	C	R	C	R	C
Experiment II	.667	.623	1.300	.860	1.667	1.860
Experiment III, no intervening task	.773	.663	1.507	1.140	1.773	2.140
Experiment III, intervening task	.687	.577	1.543	1.123	1.687	2.123

The effect of error-contingent presentation

The main concern of these experiments, of course, was with the differences between conditions R and C, and there are several points at which we might expect to find such differences. First, if the storage rate, a, is affected by error-contingent presentation, we would expect to see a difference between conditions in the number of initial errors. Second, the fact that $b \neq d$ in condition R suggests that this condition has a lower value of the parameter b, the probability that a just-stored item passes directly to state L. If this were the case, we would expect effects on both the number of initial errors and the probability that no errors follow the initial error run. Third, and most important, there are two aspects of the data from sequences with errors after the first success that relate particularly to second-stage learning. The number of errors after the first success directly reflects the probability of reaching criterion after an error. Also, the proportion of errors between the first success and last error has been shown by Greeno (1968) to be equal to $1 - (1 - c)p$, and thus this measure reflects p (performance in state H) and c (the probability of learning after successful recall).

Before detailed parametric analysis of these four aspects of the data, a brief discussion of the potential differences in terms of some summary statistics is in order. First, with regard to storage rate, the number of initial errors averaged .275 in condition R and .392 in condition C. This difference, although large, is apparently anomalous, since items in the two conditions received identical treatment until after the first success. Second, there were far fewer items with no errors after the first success in condition R than in condition C, a result which supports our suspicions of a lower value of b in the former condition; the proportions of items with no errors after the first success were .670 and .830 for conditions R and C respectively, and the difference between the two was significant [$\chi^2(1) = 20.04$, $p < .001$]. Finally, there appeared to be no differences at all between conditions after the first success. The overall precriterion error rate was .520 ($N = 250$) in condition R and .503 ($N = 503$) in condition C. The proportion of times that an error after the first success was followed by the criterion run was .777 ($N = 130$) in condition R and .753 ($N = 69$) in condition C. Thus, it appears that the major effect of error-contingent presentation is to depress the proportion, b, of items passing directly from state O to state L. The other second-stage learning parameters, c, d, and q, are apparently identical in the two conditions.

The implications of these results are best understood within the framework of a parametric analysis of the data. First, compare Figure 4, a tree

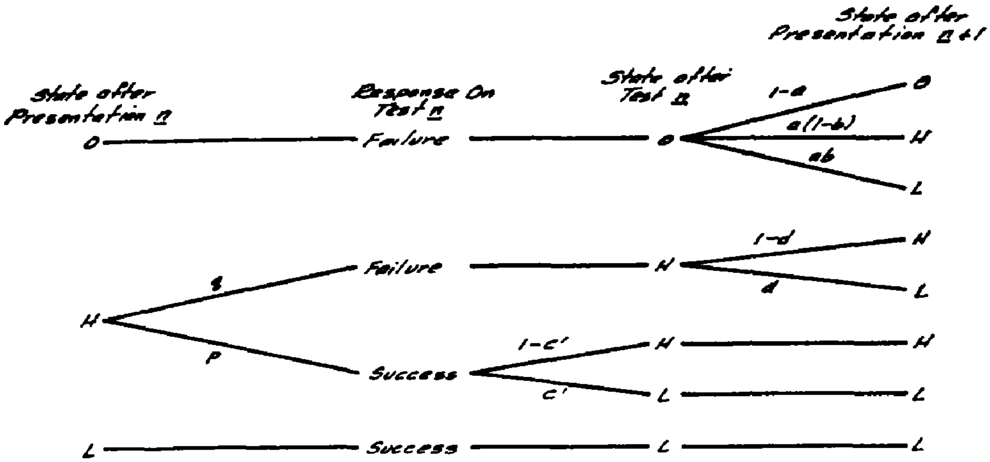


Figure 4. Tree diagram of a trial for condition R items according to the two-stage model

diagram of the process for condition R, and Figure 2, which describes the process for condition C. Note that Figure 4 was constructed by assuming that the only potential difference between the two paradigms is that d' must be zero in condition R. The rationale behind this assumption is clear. The parameter d' reflects the extent of learning on presentations following successful recall. Since such presentations did not occur in condition R, the parameter cannot be applied in this condition. In all other respects, however, conditions R and C were identical. Hence, we would not expect any of the other parameters to differ between conditions.

To apply Figure 4 to the data, recall that we can estimate only three parameters from data generated by Equation 1, namely, a $(1 - d)q$, and $(1 - c)p$; from Equation 2 we can obtain explicit expressions for the last parameter. That is, letting z_C and z_R equal $(1 - c)p$ in conditions C and R respectively,

$$(1 - c)p = \begin{cases} (1 - c')p = z_R & \text{for condition R} \\ (1 - c')(1 - d')p = z_C & \text{for condition C.} \end{cases}$$
[Equation 4]

Since error-contingent presentation can affect only d', we would expect both a and $(1 - d)q$ to be identical in both conditions, but as Equation 4 shows, z_R will exceed z_C whenever $d' > 0$, that is, whenever learning occurs on presentations following successes.

Now consider the specific implications of each of the three theories

under consideration for Equation 4. The strategy-selection theory requires that $d' = 0$ and hence that $z_R = z_C$. The law of exercise requires that $d' = d$ and hence that $z_R > z_C$. Finally, the general-potentiation theory holds that $c' = 0$ and $d' = d$. With some algebraic manipulation of Equation 4, this restriction can be translated into the requirement that $z_R = z_C/[z_C + (1 - d)q]$. To reiterate, we expect the parameters a and $(1 - d)q$ to be the same in both conditions. If, in addition, $z_R = z_C$, the strategy-selection theory would gain support. If, however, $z_R > z_C$, then one of the other two theories would be indicated, with the general-potentiation model appearing more plausible if $z_R = z_C/[z_C + (1 - d)q]$.

The technique to estimate the parameters of the model was that of Experiment I. That is, STEPIT was used to find those values of the parameters that maximized the likelihood of the data under the assumption that each of the 306 protocols was independently governed by Equation 1 under a single set of parameter values per condition. The resulting values were

$$a = .757,\ (1 - d)q = .115,\ z_R = (1 - c)p = .480 \quad \text{for condition R}$$
$$a = .543,\ (1 - d)q = .112,\ z_C = (1 - c)p = .501 \quad \text{for condition C.}$$
[Equation 5]

The ratio $z_C/[z_C + (1 - d)q]$ was estimated at .817. The difference between conditions in a and $(1 - d)q$ was not significant by Wilks's λ [$\chi^2(2) = 2.07$]. This indicates a certain amount of support for the boundary conditions of the experiment. Of more importance, z_R was slightly, but not significantly, *less* than z_C and significantly less than the ratio $z_C/[z_C + (1 - d)q]$: for the former comparison, $\chi^2(1) = .081$; for the latter, $\chi^2(1) = 75.04$, $p < .001$.

Thus, the general-potentiation model can be eliminated on the basis of these analyses, since it requires that $z_R = z_C/[z_C + (1 - d)q]$. The strategy-selection theory, with its implication that $z_R = z_C$, appears to provide the best account of the data. The law of exercise, which requires that $z_R > z_C$, cannot, however, be ruled out. Although the differences between the estimates of z_R and z_C were slightly in the wrong direction for this theory, the variance of these estimates could be large enough to make such a result likely even if the law of exercise were true. While a complete resolution of this problem is impossible, some information can be gained by putting a plausible upper bound on the parameter d'. To do this, note from Equation 4 that d' can be estimated as $1 - z_R/z_C$. It is therefore possible to determine how large d' could possibly be by testing the hypothesis that $d' = d'_0$ for various values of d'_0. Using a one-tailed Wilks's λ, values of d'_0 greater than .16 can be rejected at the .05 level; values

greater than .22 can be rejected at the .01 level. These figures, when compared to the estimate of .64 obtained for d in Experiment I, hardly support the contention that $d' = d$.

While the great bulk of the evidence points to the strategy-selection model, the difference between conditions in mean trial of the last error remains to be explained. The explanation for this effect seems to lie in the parameter b, which was abnormally low in condition R. The reason for b being low in condition R may lie in the replacement procedure used to keep presentation protocols of constant length. Recall that an item from sublist X replaced each item in condition R after the latter's first success. This introduction of new material into the list in the absence of the just-learned material may have rendered fallible some strategies that normally would have been infallible. Since the new material was introduced on the trial of the first success in condition R, this explanation is consistent with the observed low proportion of items in condition R with no errors after the first success.

The results of the experiments can be summarized quite simply. First, short-term memory for responses seemed to play no large role in learning. Second, apart from an unusually low value of b in condition R, boundary conditions for a test of the three theories were met. That is, goodness of fit to Equation 1 was good, and neither a nor $(1 - d)q$ appeared to vary across conditions. Finally, the strategy-selection theory gained support from the apparent constancy of $(1 - c)p$ in the two conditions, the general-potentiation theory provided a clearly unacceptable account of the data, and even the most plausible version of the law of exercise did not seem plausible enough.

CONCLUSION

The introduction to this paper suggested that when people memorize material over a series of trials, they spend the first few trials storing the material and further trials learning to retrieve it. The process of learning to retrieve has been the concern of this paper, and the evidence developed here strongly suggests this process is similar to that proposed by Restle (1962) for concept identification. That is, people learn to retrieve by trying out one strategy after another, rejecting each as it fails, until a reliable strategy is found. In the free-recall situation used here, strategy testing and selection occurs on each presentation, when the subject evaluates his performance on the previous trial and develops new strategies to deal with those items not recalled on the previous test.

This notion of strategy selection is difficult to relate to current theories of free recall and of list learning in general. Most of these theories are concerned more with the static organizational aspects of learning than with the dynamics of learning, which are the concern of this paper. Nonetheless, any successful theory of list learning should deal with both aspects of the process, and I will therefore conclude with some suggestions in that regard. In particular, I will first indicate how current thinking about organization in list learning corresponds to the two-stage model and then suggest how this correspondence relates to the strategy-selection theory.

Some of the best developed theories of list learning, notably those of Anderson (1972), Kintsch (1970, 1972, 1974), and Rumelhart, Lindsay, and Norman (1972), suggest that lists are remembered as collections of cues representing facts or propositions about the lexical and semantic properties of the items. These cues are of two types.

Naming cues represent the properties of individual items. They are mainly lexical in nature and allow for the recognition and reconstruction of the item. The storage stage of learning corresponds to the encoding of these naming cues. Before this step, neither recognition nor retrieval is possible. After an item's naming cue has been incorporated into memory, recognition will be virtually perfect, but retrieval will be dependent on establishing the second type of cue.

Retrieval cues represent the relations (primarily semantic) between items on the list. Given access to one part of the memory structure, these cues can guide the retrieval process to other portions. Thus, retrieval cues are the embodiment of the plan for retrieval of the list. Because they represent relations between items, their availability will be dependent on the exact sequence of retrieval operations and on the contents of primary memory. While some of the retrieval cues will be fairly stable in this regard, becoming available on virtually every trial, others will be available only when the sequence of retrieval operations happens to allow access to them. Rejecting one of the latter in favor of one of the former, more stable cues constitutes the second stage of learning, that is, learning to retrieve.

Learning to retrieve, then, involves the selection and encoding of an efficient retrieval cue, and it is through a description of this cue-selection process that the strategy-selection theory applies to list learning. Cue selection does not occur in a vacuum; rather, a cue is selected from the many other memory structures or episodes involving the item. These structures are derived not only from the list as presented but also from preexperimental experiences with the item and particularly from experi-

ences recalling the item on test trials. On each presentation, the subject will select a retrieval cue from one of these structures in such a way as to maximize likelihood that the chosen cue will be available on the subsequent test. For an item not recalled on the previous test, such a choice must be a judicious guess, but if an item was recalled on the previous test, the obvious choice of a retrieval cue is the one that supported recall on that test. This is precisely the win-stay/lose-shift policy that defines this strategy-selection theory.

In summary, this paper has presented evidence that learning only occurs on presentations following errors and thus that presentations following successes are psychologically ineffective. It has tested a promising theory of test trials in multitrial free recall and, in doing so, indicated the important role of feedback in this paradigm. And it has made a start on a theory of learning to retrieve. Compared to the enormous amount of knowledge of and concern with storage, learning to retrieve has received little attention and remains a somewhat mysterious process. The research reported in the present paper suggests that the learning-by-accumulation theories used to characterize the storage process do not well describe learning to retrieve. Rather, the dynamic, goal-directed processes characteristic of problem solving offer a more promising direction for theories of learning to retrieve.

Notes

This research was supported in part by Grant OEG-572-0016(509) from the Office of Education to the author and in part by a grant from the University Research Board of the Graduate College of the University of Illinois to the author. The author acknowledges the assistance of G. Alex Weigel and Tom Nygren in the execution and data analysis of Experiment I and that of Halff Associates of Dallas, Texas, in preparing the figures. Particular thanks are due Peter Polson of the University of Colorado for many helpful suggestions offered in a review of an earlier version. Requests for reprints should be sent to Henry M. Halff, who is now at the Office of Naval Research, Arlington, Virginia 22217.

1. For brevity's sake, we will refer to each of these classes of theories as a 'theory.' The reader should keep in mind that the defining characteristics of each class are broad enough to accommodate several specific theories.

2. The astute reader will note that the test used here is not very powerful because of the small sample sizes. To illustrate this, we evaluated χ^2 (not corrected for continuity) using specific values of c' to generate expected values. The best-fitting value of c' is .167, but only values that exceed .560 are significant at the .05 level, and a value of .622 is required for rejection at the .01 level.

3. The trial of the first success was eliminated from the analysis due to the above-mentioned discrepancy of fit on that trial. When it is included, estimates

of y and f are 79/126 and 1.219 respectively. The latter figure is not significantly different from 1 [$\chi^2(1) = .35$].

4. Using the same techniques as described in note 2, we found that values of f lower than .514 are rejected at the .05 level and those lower than .433 are rejected at the .01 level.

References

Adams, J. A., and Bray, N. W. 1970. A closed-loop theory of paired-associate verbal learning. *Psychological Review* 77:385–405.

Anderson, J. R. 1972. FRAN, a simulation model of free recall. In *The psychology of learning and motivation: Advances in research and theory,* vol. 5, ed. G. H. Bower. New York: Academic Press.

Birnbaum, I. M., and Eichner, J. T. 1971. Study versus test trials and long-term retention in free-recall learning. *Journal of Verbal Learning and Verbal Behavior* 10:516–521.

Bjork, R. A. 1975. Retrieval as a memory modifier: An interpretation of negative recency and related phenomena. In *Information processing and cognition: The Loyola symposium,* ed. R. L. Solso. Hillsdale, N.J.: Erlbaum Associates.

Bower, G. H. 1970. Organizational factors in memory. *Cognitive Psychology* 1:18–46.

Buschke, H. 1973. Selective reminding for analysis of memory and learning. *Journal of Verbal Learning and Verbal Behavior* 12:543–550.

Buschke, H. 1974. Retrieval in verbal learning. *Transactions of the New York Academy of Sciences,* series II, 36:721–729.

Buschke, H. 1975. Overcoming difficulties in the experimental analysis of learning. Paper presented at the Eighth Indiana Theoretical and Mathematical Psychology conference in Bloomington in April.

Chandler, J. P. 1969. STEPIT — finds local minimum of a smooth function of several parameters. *Behavioral Science* 14:81–82. CPA 312.

Donaldson, W. 1971. Output effects in multitrial free recall. *Journal of Verbal Learning and Verbal Behavior* 10:577–585.

Greeno, J. G. 1968. Identifiability and statistical properties of two-stage learning with no successes in the initial stage. *Psychometrika* 33:173–215.

Greeno, J. G. 1970. How associations are memorized. In *Models of human memory,* ed. D. A. Norman: New York: Academic Press.

Greeno, J. G., James, C. T., and DaPolito, F. J. 1971. A cognitive interpretation of negative transfer and forgetting of paired associates. *Journal of Verbal Learning and Verbal Behavior* 10:331–345.

Greeno, J. G., and Scandura, J. M. 1966. All-or-none transfer based on verbally mediated concepts. *Journal of Mathematical Psychology* 3:388–411.

Halff, H. M. 1976. Parameterizations of Markov models for two-stage learning. *Journal of Mathematical Psychology* 14:123–129.

Heine, R. T. 1970. Quantitative analysis of multitrial free-recall learning. University of Michigan Mathematical Psychology Program, technical report MMPP 70-12.

Hilgard, E. R., and Bower, G. H. 1975. *Theories of learning.* 4th ed. Englewood Cliffs, N.J.: Prentice-Hall.

Hogan, R. M., and Kintsch, W. 1971. Differential effects of study and test trials on long-term retention and recall. *Journal of Verbal Learning and Verbal Behavior* 10 562–567.

Humphreys, M., and Greno, J. G. 1970. Interpretation of the two-stage analysis of paired-associate memorizing. *Journal of Mathematical Psychology* 7:275–292.

Izawa, C. 1971. The test trial potentiating model. *Journal of Mathematical Psychology* 8:200–224.

Kintsch, W. 1970. Models for free recall and recognition. In *Models of human memory*, ed. D. A. Norman. New York: Academic Press.

Kintsch, W. 1972. Notes on the structure of semantic memory. In *Organization of memory*, ed. E. Tulving and W. Donaldson. New York: Academic Press.

Kintsch, W. 1974. *The representation of meaning in memory*. Hillsdale, N.J.: Erlbaum Associates.

Kintsch, W., and Morris, C. J. 1965. Application of a Markov model to free recall and recognition. *Journal of Experimental Psychology* 69:200–206.

Lachman, R., and Laughery, K. R. 1968. Is a test trial a training trial in free-recall learning? *Journal of Experimental Psychology* 76:40–50.

Murdock, B. B., Jr., Anderson, R. E., and Ho, E. 1974. Effects of presentation order on learning in multitrial free recall. *Journal of Verbal Learning and Verbal Behavior* 13:522–529.

Murdock, B. B., Jr., Penney, C., and Aamiry, A. 1970. Interactive presentation in multitrial free recall. *Journal of Verbal Learning and Verbal Behavior* 9:679–683.

Pagel, J. C. 1973. A Markov analysis of transfer in paired-associate learning with high intralist similarity. *Journal of Verbal Learning and Verbal Behavior* 12:456–470.

Polson, M. C., Restle, F., and Polson, P. G. 1965. Association and discrimination in paired-associates learning. *Journal of Experimental Psychology* 69:47–55.

Polson, P. G. 1972. A quantitative theory of concept identification in the Hull paradigm. *Journal of Mathematical Psychology* 9:141–167.

Restle, F. 1962. The selection of strategies in cue learning. *Psychological Review* 69:329–343.

Restle, F. 1964. Sources of difficulty in learning paired associates. In *Studies in mathematical psychology*, ed. R. C. Atkinson. Stanford, Calif.: Stanford University Press.

Restle, F. 1965. The significance of all-or-none learning. *Psychological Bulletin* 62:313–324.

Rosner, S. R. 1970. The effects of presentation and recall trials on organization in multitrial free recall. *Journal of Verbal Learning and Verbal Behavior* 9:69–74.

Rumelhart, D. E., Lindsay, P. H., and Norman, D. A. 1972. A process model for long-term memory. In *Organization of memory*, ed. E. Tulving and W. Donaldson. New York: Academic Press.

Shiffrin, R. M. 1970. Memory search. In *Models of human memory*, ed. D. A. Norman. New York: Academic Press.

Thompson, D. M., and Tulving, E. 1970. Associative encoding and retrieval: Weak and strong cues. *Journal of Experimental Psychology* 86:255–262.

Tulving, E. 1962. Subjective organization in free recall of "unrelated" words. *Psychological Review* 69:344–354.

Tulving, E. 1967. The effects of presentation and recall of material in free-recall learning. *Journal of Verbal Learning and Verbal Behavior* 6:175–184.

Tulving, E., and Osler, S. 1968. Effectiveness of retrieval cues in memory for words. *Journal of Experimental Psychology* 77:593–601.

Tulving, E., and Thompson, D. M. 1973. Encoding specificity and retrieval processes in episodic memory. *Psychological Review* 80:352–373.

Waugh, N. C., and Smith, J. E. 1962. A stochastic model for free recall. *Psychometrika* 27:141–154.

Wilks, S. S. 1962. *Mathematical statistics.* New York: Wiley.

AMERICAN JOURNAL OF PSYCHOLOGY
September 1977, Vol. 90, No. 3, pp. 407–418

Long-term effects of prior experience in attenuating amnesia

Charles F. Hinderliter and David C. Riccio
Kent State University

In Experiment I, prior experience with passive-avoidance training followed by latent extinction was given 1, 3, 5, or 15 days before criterion (re)training and an amnesic treatment. It produced nearly complete protection from retrograde amnesia at the three shorter intervals; at the longest interval, amnesia was present but less severe than in a control group without the familiarization. In Experiment II, prior experience was given 1, 5, or 15 days before a noncontingent shock and an amnesic treatment. Evidence of a reactivation of memory was obtained only at the longest interval. Thus, familiarization and reactivation seem to represent different processes. The results are interpreted as consistent with explanations stressing the disruption of retrieval in retrograde amnesia.

Retrograde amnesia is typically produced when electroconvulsive shock (DeVietti and Hopfer, 1974), hypothermia (Beitel and Porter, 1968; Riccio, Hodges, and Randall, 1968), or other agents (McGaugh and Herz, 1972) are delivered to an animal within a relatively short period after learning. Despite the reliability of this phenomenon, there are also a variety of behavioral manipulations which, when given as pretraining or *prior experience,* attenuate or eliminate the amnesic effects of these agents. Allowing a subject to become familiarized with the training apparatus (Lewis, Miller, and Misanin, 1968, 1969), with a sequence of training and amnesic treatments (Kesner, McDonough, and Doty, 1970; Nachman and Meinecke, 1969; Riccio and Stikes, 1969), with the amnesic agent or the aversive stimulus to be used in training (Hinderliter, Smith, and Misanin, 1973; or a placebo injection, David-Remacle, 1973), with training that is followed by extinction (Jensen and Riccio, 1970), or with training on related tasks (Jensen, Riccio, and Gehres, 1975) all can modulate the extent of memory loss.

While the generality of an effect of prior experience seems well established, systematic information about its durability or persistence is lacking.

It would seem important to know whether prior experience exerts only a transient influence in blocking amnesia or whether it has that effect over relatively long intervals of time. In a different context, Leaton (1976) recently posed a similar question about habituation. He found that a single stimulus exposure could produce a surprisingly long-term modification of responsivity.

The purpose of the present study was to see whether a latent-extinction procedure (previously used by Jensen and Riccio, 1970) would be effective in attenuating amnesia if it preceded the training and amnesic treatments by 1, 3, 5, or 15 days. In essence then, the empirical question asked was: How long after the 'memory' of a prior experience had been established would amnesia produced by hypothermia be attenuated or prevented?

EXPERIMENT I

METHOD

Subjects

The subjects were 117 Holtzman male albino rats of 278–430 g. They had been caged in groups of 4 to 6 for at least a week after receipt from the supplier before the start of the experiment. Rats were maintained on ad lib food and water at all times.

Apparatus

The passive-avoidance apparatus consisted of a 38.1-by-16.5-by-23.5-cm Plexiglas box divided into two equal chambers by a partition with a 7.5-by-6.4-cm opening. A wooden door was placed inside the partition to block the opening when necessary. One chamber of the apparatus was white with a clear Plexiglas lid and a solid metal floor. The other chamber was black with a black Plexiglas lid and a grid floor consisting of .25-cm stainless steel rods spaced 1.25 cm apart. A Foringer model SC-901 scrambler and a matched-impedance ac source of shock (Campbell and Teghtsoonian, 1958) were used to deliver a 150-V shock through the grids. The room containing the passive-avoidance apparatus was illuminated by a 15-W bulb suspended 38 cm above the white chamber. A model 2095 Forma Temp. Jr. bath and circulator containing tap water maintained at 3–5 deg C was used to produce hypothermia. Rectal temperatures were taken with a model 43TD Yellow Springs Tele-Thermometer inserted approximately 3.2 cm into the anus.

Procedure

The general plan of this experiment consisted of three phases. Phase 1 involved pretraining the 108 experimental animals on a passive-avoidance task and then extinguishing this response. This procedure was chosen as the prior experience because it has been shown to be quite effective in preventing amnesia (Jensen and Riccio, 1970). In phase 2, the rats were (re)trained and/or underwent

hypothermia 1, 3, 5, or 15 days after phase 1. Retention of the passive-avoidance task was examined in phase 3.

More specifically, 24 hr before phase 1, each rat was handled for 2 to 3 min, weighed, had an ear punched for the purpose of identification, and placed in an individual cage. In phase 1, all of the experimental animals received passive-avoidance pretraining. Each subject was placed in the white side of the apparatus facing away from the door. The door was opened 5 sec later. Upon placing all four feet in the black chamber, each rat received a 1-sec, 150-V inescapable foot shock. The time it took the rat to cross into the black chamber was recorded to the nearest .1 sec as the measure of latency. Each rat was then returned to its home cage and 3 hr later underwent the first of two 15-min extinction sessions. In these sessions, rats were placed in the white chamber for 5 min, then in the black chamber for 10 min. Rats could not cross between chambers, as the partition door was down at all times. The second extinction session, which occurred 3 hr after the first extinction session, ended phase 1. The particular parameters chosen for the extinction session were similar to those of a previous study (Jensen and Riccio, 1970).

In phase 2, each experimental animal was assigned to one of twelve groups receiving footshock (FS) and/or hypothermia (H) 1, 3, 5, or 15 days after phase 1 had been completed. The rats in the four FS-H groups were given passive-avoidance training with the footshock in phase 2 and latency to enter the black chamber was recorded. Within 45 sec after that (re)training, these rats were restrained in wire mesh cylinders and immersed into water at 3–5 deg C. If rectal temperatures were above 21 deg C after 10.5 min, rats were reimmersed an additional 30–60 sec and temperatures were checked again, this procedure being repeated until temperatures fell at or below 21 deg C. Rats then were dried with paper towels and returned to their home cages. No rat was immersed longer than 12 min. The rats in the four FS-nH groups were given passive-avoidance training with the footshock in phase 2 but were then returned to their home cages. This group was used to determine whether varying the amount of time between the prior experience and the (re)training differentially affected 24-hr retention of the (re)trained response. The rats in the four nFS-H groups were treated like those in the FS-H groups in phase 2 except that no footshock was given when rats crossed into the black chamber. These nFS-H groups were used to assess any differences in performance or aversive effects attributable to hypothermia.

In order to examine the effects of prior experience on preventing amnesia, it is necessary to show that rats not receiving prior experience would show evidence of amnesia. This was achieved by running as controls rats that received the same treatment as the FS-H groups except that the control rats did not have the prior experience with the passive-avoidance task and extinction in phase 1. In an attempt to examine any effects different amounts of isolation might have on the amnesic effects of hypothermia, the control animals received the training and hypothermia treatments either 2 ($N = 2$), 4 ($N = 3$), 6 ($N = 2$), or 16 ($N = 2$) days after handling and isolation.

In phase 3, conducted 24 hr after phase 2, there was a passive-avoidance test for all animals. Each subject was placed in the white chamber and the time it took the rat to cross into the black chamber was recorded. Rats not stepping through within 10 min were arbitrarily assigned a test latency of 600 sec.

During the experiment, 13 rats died and were replaced. No more than 2 rats from any one group died. In phase 2, there were 2 rats that failed to cross into the black chamber after 5 min and were also replaced.

RESULTS

Latencies in phase 1 and phase 2

The initial latencies in phase 1 were brief, with means of 16.3, 12.4, and 12.2 sec for the FS-H, FS-nH, and nFS-H groups respectively. Response times were slightly longer in phase 2: collapsing across interphase intervals for the treatment groups, the mean latencies were 21.1, 22.8, and 32.2 sec for groups FS-H, FS-nH, and nFS-H. (The control animals received no prior experience; their mean initial latency at training was 7.9 sec.)

A 3 × 4 × 2 factorial analysis of variance with one repeated measure was performed on the latencies obtained in phases 1 and 2 for the twelve groups receiving prior experience. The only significant effect was for phase [$F(1, 9) = 10.86, p < .002$], suggesting either that extinction had not been complete or that there was some spontaneous recovery that reached asymptote at 24 hr after extinction. Although latencies obtained in phase 2 were slightly longer than those obtained in phase 1, comparisons with other data from this laboratory indicate that the extinction treatment in phase 1 was highly effective. For example, groups with similar footshock training *not* followed by extinction and given retention tests from 1 to 10 days later had median latencies ranging between 485 and 600 sec (Jensen and Riccio, 1970; Riccio and Stikes, 1969), whereas 27 sec was the longest median latency of all groups in phase 2. It should also be noted that a 3 × 4 factorial analysis of variance with one added control (Winer, 1962, pp. 263–267) performed on the phase 2 data on latency, including that for the control animals, revealed no significant differences between any of the groups. Thus, this particular measure of retention also indicates that the groups performed comparably prior to the footshock and/or hypothermia in phase 2.

Latencies in phase 3

Table 1 shows the median latencies in phase 3 (test latencies) for all groups. Because 25% of the rats had scores at ceiling, nonparametric two-tailed tests were used in making all comparisons. As would be expected, the passive-avoidance latencies of the control animals were significantly lower than those of the FS-nH groups at all intervals [Us$\leq$ 10,

Table 1. Median test latencies (sec) for all groups; phase 3, Experiment I

	Retention interval between phases 1 and 2 (in days)			
	1	3	5	15
FS–nH	600.0	342.4	600.0	600.0
FS–H	139.9	100.1	222.9	43.8
nFS–H	14 6	5.1	2.3	8.7
Controls		8.8		

Note: Data for controls are entered for purposes of comparison; these animals did not in fact have a phase 1, phase 2 being their first encounter with the passive-avoidance task as well as with the hypothermia.

n_1/n_2s $= 9$, ps $\leq .02$] but did not differ from those of the four nFS-H groups. The severity of amnesia is reflected in the fact that rats in the control groups showed no significant increase in latencies between phases 2 and 3. Because rats given only one session of footshock training have been run as a control for retention so frequently in our laboratory, that condition was not included here. However, it may be instructive to note that median scores under that condition have typically been well over 400 sec, in contrast with the median under 9 sec that was obtained for the controls in the present experiment.

The data of primary interest concern the effect of the familiarization or pretraining on the magnitude of amnesia. At each interval between phases, a Kruskal-Wallis analysis indicated significant treatment effects among groups, including the controls [$H = 15.3$, $df = 3$, $p < .01$ for 1-day groups; $H = 19.3$, $df = 3$, ps $< .001$ for 3-, 5-, and 15-day groups]. More detailed comparisons indicated that prior experience attenuated or eliminated amnesia at the 1-, 3-, and 5-day intervals: latencies of the three FS-H groups were significantly longer than those of either the controls or the respective nFS-H groups [Us ≤ 10, $n_1/n_2 = 9$, ps $< .02$] but did not differ from those of the respective FS-nH groups [Us ≥ 19, $n_1/n_2 = 9$, ps $> .05$]. These results replicate and extend the previously reported effectiveness of pretraining followed by extinction in reducing amnesia when (re)training and hypothermia treatments follow 24 hr later (Jensen and Riccio, 1970). Results obtained at the 15-day interval suggest that this protection from amnesia decreases in effectiveness if the interval between the prior experience and the training with the amnesic treatment is sufficiently long. In contrast with the findings at the shorter intervals, at the 15-day interval the latencies of the FS-H group were significantly less than those of the FS-nH group [$U = 11$, $n_1/n_2 = 9$, $p < .02$] and not

different from the nFS-H group [$U = 19, n_1/n_2 = 9, p > .05$]. Even at 15 days, however, amnesia was attenuated in the experimental groups, as indicated by the finding that latencies of the FS-H group were longer than those of the controls [$U = 5.5, n_1/n_2 = 9, p < .002$]. Thus, pretraining and extinction effectively prevented amnesia 5 days after it had been administered and was still partially effective 15 days later. This persistence of prior experience in attenuating amnesia indicates that motivational variables or changes in responsiveness to the amnesic agent (see Chorover and DeLuca, 1969) are probably not responsible for the effects of prior experience.

The short median latencies of the nFS-H groups at the 1-, 3-, 5-, and 15-day intervals suggest that any common arousal or stress-eliciting properties shared by footshock and hypothermia were not sufficient to remind rats of the punishment contingency that had been established and extinguished in phase 1. The results obtained for the nFS-H groups also indicated that the long latencies of the FS-H groups do not simply reflect artifacts attributable to the combined effects of aversive pretraining (phase 1) and hypothermia (phase 2).

Varying the interval between the prior experience and (re)training did not differentially affect 24-hr retention of the (re)trained passive-avoidance response. Latencies of the FS-nH groups did not differ across retention intervals but were significantly higher than those of their respective nFS-H groups at each interval [$Us \leq 13, n_1/n_2 = 9, ps < .02$]. Also, of all the FS-nH groups, the highest proportion of maximum scores (600 sec) occurred at the 15-day interval. Thus, there is no evidence from the performance of the groups that received only the footshock training in phase 2 (groups FS-nH) to suggest that the degree of protection from amnesia is attributable to different levels of learning in groups before the amnesic treatment.

DISCUSSION

The present findings are consistent with a number of other studies in showing that prior experience can markedly diminish the extent of memory loss induced by amnesic agents (e.g., Miller, 1970; David-Remacle, 1973). In addition, this experiment demonstrates that the effect of prior experience is not merely a transient or momentary modification of the organism. Within the context of the present parameters, a simple behavioral manipulation produced a strong and relatively enduring attenuation of the usual hypothermia-induced retrograde amnesia. Familiarization by pretraining and extinction greatly attenuated, or possibly prevented, am-

nesia for at least 5 days and reduced amnesia even after 15 days. At the same time, the intermediate status of the data for day 15 suggests that something akin to a decrement in retention of the pretraining or prior experience may occur.

Although this study was not designed to test particular explanations of retrograde amnesia, the outcome seems generally consistent with the interpretations that stress retrieval. For example, Lewis et al. (1969) have suggested that pretraining familiarization may provide an opportunity for attributes to be 'tagged' and elaborated into a more complete organization when criterion training occurs, thus reducing the effectiveness of an amnesic treatment (see Lewis, 1976). The data also appear to be consistent with our 'contextual cues' analysis, in which we suggested that hypothermia might serve as part of the encoding of the target memory (Hinderliter, Webster, and Riccio, 1975). Amnesia presumably occurs because the contextual cues associated with the training and amnesic treatment are not available to permit memory retrieval at testing (see Spear, 1973; Tulving, 1974). Conversely, maintaining or restoring cues (e.g., by testing shortly after treatment or by recooling subjects before testing; Hinderliter et al., 1975) tends to eliminate amnesia.

Our interpretation implies that prior experience attenuates amnesia by degrading the role of the amnesic treatment as a contextual cue. Thus, manipulations that differentiate training and hypothermia treatments should alter the amnesic effects of hypothermia. For instance, the saliency of hypothermia as a contextual cue could be reduced either by having the subjects experience beforehand either aspects of the training situation not associated with hypothermia or hypothermia in situations not associated with training. Current data indicate that prior experience with deep body cooling does prevent subsequent hypothermia-induced amnesia, as in the case of familiarization with the apparatus (e.g., Miller, 1970); however, more than one administration is required to be effective (Hinderliter and Riccio, 1976). Moreover, if differentiation (discrimination) of the training situation and amnesic treatment is produced through learning processes, then like other learning events, it might be expected to weaken over time. The present finding of diminished protection from retrograde amnesia as the time between the prior experience and the training and amnesic treatments increased is in accord with this expectation.

EXPERIMENT II

In Experiment I, it was assumed that the long latencies of the experimental groups reflected the reduced effectiveness of an amnesic treatment

in impairing recent learning. While the performance of the nFS-H groups indicated that increased latencies were not attributable to debilitation from cooling per se, it remained possible that a combination of footshock and cooling might act synergistically to modify responding. Previous evidence from our lab indicated little if any increase in the latencies of untrained rats 24 hr after noncontingent footshock and cooling (Hinderliter et al., 1975). However, the present design introduces the possibility of associative consequences from such treatment. For example, noncontingent footshock and hypothermia might serve to reactivate the fear response that had been extinguished. The relevance of this concern is highlighted by a report that appeared after Experiment I was completed. Rescorla and Heth (1975) found that after extinction of a conditioned emotional response, suppression to the conditioned stimulus could be reinstated by shocks given outside of the training apparatus. On the other hand, any such reactivation here might also be obliterated by deep body cooling.

Experiment II was conducted to determine whether a fear-motivated response is reactivated after extinction by exposure to noncontingent footshock and hypothermia. The retention intervals between extinction and treatment were the same as in Experiment I except that the 3-day groups were omitted.

METHOD

Subjects

The subjects were 27 Holtzman male albino rats around 75–100 days old. The apparatus was the same as that used in Experiment I. In addition, an 'open field,' a 34.5-by-34-by-27-cm chamber mounted on a grid floor consisting of .15-cm steel rods spaced .5 cm apart, was used to administer noncontingent footshock. To enhance its distinctiveness from the passive-avoidance chamber, cardboard inserts (manilla folder material) lined its walls.

The procedure for phase 1 (passive-avoidance training followed by two extinction sessions) was identical to that of Experiment I. In phase 2, separate groups of rats received noncontingent footshock, identical in intensity to the training shock of Experiment I, either 1 ($N = 10$), 5 ($N = 7$), or 15 ($N = 10$) days after familiarization. All rats underwent hypothermia immediately after their noncontingent shock. Then, 24 hr later they were placed in the passive-avoidance chamber and latency to cross into the previously shocked side was recorded.

RESULTS AND DISCUSSION

Median test latencies were 27, 106, and 231 sec for the 1-, 5-, and 15-

day groups respectively. Kruskal-Wallis one-way analysis of variance revealed significant effects of the retention interval [$H = 7.07$, $df = 2$, $p < .05$]. Individual comparisons revealed that the latencies of the 15-day group were longer than those of the 1- and 5-day groups [Us ≤ 19, $n_1 = 7$ or 10, $n_2 = 10$, ps $< .02$]. Although the 5-day group appeared to be intermediate, it did not differ significantly from the 1-day group. Because the data here were obtained after completion of Experiment I, direct statistical comparison with those groups cannot be made. However, the critical feature of the data is that latencies were affected by the 'control' manipulation. Although administration of noncontingent shock and hypothermia 1 day after latent extinction of a passive-avoidance response had little effect on test latencies, a different pattern emerged with longer intervals between the extinction in phase 1 and the treament in phase 2. At 15 days, there appears to have been some reinstatement of the originally acquired behavior. As the interval between treatment and testing was held constant, the changes in latencies cannot be due to differential motor impairment or other performance artifacts. Moreover, rats receiving a similar treatment without pretraining typically show very brief latencies (Hinderliter et al., 1975). Even casual comparison with Experiment I suggests that spontaneous recovery over time is not involved, as latencies at time of (re)training in phase 2 were well under a minute at all retention intervals. Apparently, memory for the previously extinguished response can be reinstated by some aspects of our treatment. This finding is generally consistent with the report that a conditioned emotional response that has been extinguished is reinduced by noncontingent footshock (Rescorla and Heth, 1975).

The increase in reactivation with longer retention intervals between extinction and treatment, while unexpected, may reflect changes in the discriminability of the two events. That is, rats receiving a footshock in a different apparatus shortly after acquisition and extinction may differentiate between the two situations, resulting in poor reactivation. If, with longer intervals, generalization gradients tend to flatten (McAllister and McAllister, 1963; Perkins and Weyant, 1958), the noncontingent footshock may not be clearly discriminated from the training episode. Under these conditions, reactivation of memory of a prior aversive event may well occur. Whether the hypothermia portion of the reminder contributes to the cues, or has some amnesic effect on the reactivated memory, cannot be ascertained in the absence of a control group receiving only the noncontingent shock. The data from Experiment I gave no indication, however, that deep cooling per se produces reactivation.

GENERAL DISCUSSION

In Experiment I, the effectiveness of an amnesic agent was markedly reduced when pretraining and latent extinction preceded the criterion task, and this protection from amnesia persisted to some degree for two weeks. In Experiment II, intended as a procedural control, memory for the pretraining was apparently reactivated when noncontingent footshock and hypothermia were given. Although the latter finding means that caution must be exercised in interpreting certain aspects of the familiarization data, we think the outcomes reflect the effect of two different processes. It is important to note that the functional relationships obtained in the two experiments were in opposite directions: at the short retention intervals, resistance to amnesia was strong and the reactivation effect was weak or intermediate; at the 15-day interval, protection from amnesia had diminished but reactivation was quite potent. Thus, the conclusions reached in Experiment I on the persistence of the effect of familiarization still appear warranted. At the same time, the data on reactivation point to a phenomenon deserving more systematic examination in its own right.

The procedure for pretraining and extinction used here was chosen as the familiarization experience because of its potency in attenuating amnesia at short intervals (Jensen and Riccio, 1970). It will now be of interest to determine whether other forms of prior experience, such as familiarization with the training apparatus or with components of the amnesic treatment, also produce persistent but declining protection from amnesia. Because these other manipulations do not involve prior training, some of the methodological complications inherent in the present design will be circumvented.

Notes

This investigation was supported in part by NSF Grant GB-41488 to the second author. This report is based on a thesis submitted to Kent State University in partial fulfillment of the requirements for the M.A. Requests for offprints should be sent to David C. Riccio, Department of Psychology, Kent State University, Kent, Ohio 44242. Received for publication May, 1976; revision, October 5, 1976.

References

Beitel, R. E., and Porter, P. B. 1968. Deficits in retention and impairments in learning induced by severe hypothermia in mice. *Journal of Comparative and Physiological Psychology* 66:53–59.

Campbell, B. A., and Teghtsoonian, R. 1958. Electrical and behavioral effects of different types of shock stimuli on the rat. *Journal of Comparative and Physiological Psychology* 51:185–192.

Chorover, S. L., and DeLuca, A. M. 1969. Transient change in electrocorticographic reaction to ECS in the rat following footshock. *Journal of Comparative and Physiological Psychology* 69:141–149.

David-Remacle, M. 1973. Attenuation of anoxia induced retrograde amnesia in rats by a pretraining placebo injection. *Physiology and Behavior* 10:693–696.

DeVietti, T. L., and Hopfer, T. M. 1974. Reinstatement of memory rats: Dependence upon two forms of retrieval deficit following electroconvulsive shock. *Journal of Comparative and Physiological Psychology* 88:1090–1099.

Hinderliter, C. F., and Riccio, D. C. 1976. Retrograde amnesia: State-dependency and contextual cues. Paper presented at Midwestern Psychological Association, Chicago.

Hinderliter, C. F., Smith, S. L., and Misanin, J. R. 1973. Effect of pretraining experience on retention of a passive avoidance task following ECS. *Physiology and Behavior* 10:671–675.

Hinderliter, C. F., Webster, T., and Riccio, D. C. 1975. Amnesia induced by hypothermia as a function of treatment test interval and recooling in rats. *Animal Learning and Behavior* 3:257–263.

Jensen, R. A., and Riccio, D. C. 1970. Effects of prior experience upon retrograde amnesia produced by hypothermia. *Physiology and Behavior* 5:1291–1294.

Jensen, R. A., Riccio, D. C., and Gehres, L. 1975. Effects of prior aversive experience upon retrograde amnesia induced by hypothermia. *Physiology and Behavior* 15:165–169.

Kesner, R. P., McDonough, J. H., Jr., and Doty, R. W. 1970. Diminished amnesic effect of a second electroconvulsive seizure. *Experimental Neurology* 27:527–533.

Leaton, R. N. 1976. Long-term retention of the habituation of lick suppression and startle response produced by a single auditory stimulus. *Journal of Experimental Psychology: Animal Behavior Processes* 2:248–259.

Lewis, D. J. 1976. A cognitive approach to experimental amnesia. *American Journal of Psychology* 89:51–80.

Lewis, D. J., Miller, R. R., and Misanin, J. R. 1968. Control of retrograde amnesia. *Journal of Comparative and Physiological Psychology* 66:48–52.

Lewis, D. J., Miller, R. R., and Misanin, J. R. 1969. Selective amnesia in rats produced by electroconvulsive shock. *Journal of Comparative and Physiological Psychology* 69:136–140.

McAllister, W. R., and McAllister, D. E. 1963. Increase over time in the stimulus generalization of acquired fear. *Journal of Experimental Psychology* 65:576–582.

McGaugh, J. L., and Herz, M. J. 1972. *Memory consolidation.* San Francisco: Albion Publishing.

Miller, R. R. 1970. Effects of environmental complexity on amnesia induced by electroconvulsive shock. *Journal of Comparative and Physiological Psychology* 71:267–275.

Nachman, M., and Meinecke, R. O. 1969. Lack of retrograde amnesia effects

of repeated electroconvulsive shock and carbon dioxide treatments. *Journal of Comparative and Physiological Psychology* 68:631–636.

Perkins, C. C., Jr., and Weyant, R. G. 1958. Interval between training and test as a determinant of the slope of generalization gradients. *Journal of Comparative and Physiological Psychology* 51:596–600.

Rescorla, R. A., and Heth, C. D. 1975. Reinstatement of fear to an extinguished conditioned stimulus. *Journal of Experimental Psychology: Animal Behavior Processes* 104:88–96.

Riccio, D. C., Hodges, L. A., and Randall, P. K. 1968. Retrograde amnesia produced by hypothermia in rats. *Journal of Comparative and Physiological Psychology* 66:618–622.

Riccio, D. C., and Stikes, E. R. 1969. Persistent but modifiable retrograde amnesia produced by hypothermia. *Physiology and Behavior* 4:649–652.

Spear, N. E. 1973. Retrieval of memory in animals. *Psychological Review* 80: 163–194.

Tulving, E. 1974. Cue-dependent forgetting. *American Scientist* 62:74–82.

Winer, B. J. 1962. *Statistical principles in experimental design.* New York: McGraw-Hill.

AMERICAN JOURNAL OF PSYCHOLOGY
September 1977, Vol. 90, No. 3, pp. 419–429

List discrimination during transfer in free recall

James W. Pellegrino and Judith Petrich
University of Pittsburgh

The free recall of successive partially overlapping lists was studied, using Anderson and Bower's 1972 model to derive specific predictions. All predictions were supported by the data. Specifically, items repeated across lists showed depressed recall relative to new items, particularly when repeated across multiple lists. Knowledge of the interlist relationship served only to reduce the level of negative transfer relative to an uninformed condition. The data on a test of list identification given to the experimental informed and uninformed conditions and to a control condition with nonoverlapping lists also support the model of the role of list tagging and contextual elements in the recall decision.

One model of list discrimination in learning for free recall involves *list tagging,* a process whereby unique tags are associated to individual items in a particular list. These tags and their associated *contextual elements* serve to identify an item's membership in one or more lists (Anderson and Bower, 1972). As an item is presented in successive lists, the process of establishing new tags becomes increasingly more difficult, an effect similar to negative transfer under an A-B A-C paradigm. Such a model is consistent with Anderson and Bower's report (1972) of a curvilinear function for free recall over successive single-trial lists with overlapping items. The difficulty of establishing new tags is also hypothesized to contribute to negative transfer in part/whole and whole/part free recall. Consistent with this is the localization of negative part/whole transfer in the old items (e.g., Bower and Lesgold, 1969).

Sternberg and Bower (1974) have argued that the model is also capable of explaining the elimination of negative transfer and the occurrence of positive transfer when the subject is informed of the interlist relationship (e.g., Novinski, 1972). Their explanation of the latter was that knowledge of the interlist relationship provides a rule that eliminates the necessity for retagging old items. Petrich, Pellegrino, and Dhawan (1975) showed that the model is also capable of explaining additional phenomena of

part/whole transfer if it is expanded to include a strategy giving priority to new items (see also Schulze and Gorfein, 1976). That is, knowledge that all list 1 items are also in list 2 allows the learner to focus attention on the new items (Petrich et al., 1975) while simultaneously establishing the relevance of list 1 tags (Sternberg and Bower, 1974).

Petrich et al. (1975) tested the generality of the model by using two partially overlapping lists whereby half of the list 1 items were included in list 2 along with an equal number of new items. This design was examined under the typical conditions and also under conditions where the subject was fully informed of the interlist relationship. Unlike the part/whole design, knowledge of the interlist relationship in this design should not produce positive transfer, since retagging is still required for the old items. The results of Petrich et al. (1975) supported this prediction, although knowledge of the interlist relationship facilitated performance relative to an uninformed condition.

The present research attempted to replicate and extend the results of Petrich et al. (1975) to three partially overlapping lists. As in the study by Petrich et al. (1975), list 2 was composed of half of list 1 together with an equal number of new items. List 3 was composed of half of list 2, together with an equal number of new items. However, the old items in list 3 were equally divided between items that had appeared only in list 2 and items that had appeared both in lists 1 and 2. Anderson and Bower's model explicitly predicts that the latter items should show the poorest recall, because of the increased difficulty of establishing new tags for items that already have two prior tags as opposed to one or no prior tags. Knowledge of the interlist relationships should not produce positive transfer, relative to a control condition, during either list 2 or list 3 learning.

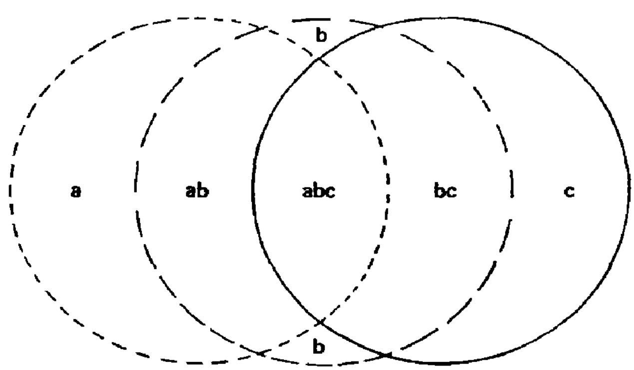
Petrich et al. (1975) also collected data on a test of list identification that followed free recall. The data were consistent with Anderson and Bower's model. This was the case for overall levels of correct list identification as well as the distributions of error for different item types (i.e., items from different lists).

The model can also be used to predict performance in the present design. Figure 1 provides a basis for deriving predictions for each of the various item types. As represented in this figure, there is a pool of contextual elements active during the learning of each list. These elements become associated to the tag for each item and form the basis for list identification. Since the successive lists involve overlapping items, then the lists will also share elements as shown in Figure 1. Thus, items appearing only in list 1 will have elements unique to list 1 (a), common to lists 1 + 2 (ab), and common to all three lists (abc). Items appearing only in

list 2 will have elements unique to list 2 (b), common to lists 1 + 2 (ab), common to lists 2 + 3 (bc), and common to all three lists (abc). Items appearing only in list 3 will have elements unique to list 3 (c), common to lists 2 + 3 (bc), and common to all three lists (abc). The element types for items appearing in more than one list — in lists 1 + 2, 2 + 3, or 1 + 2 + 3 — are also shown in Figure 1.

If one assumes a list-dominance factor, then elements common to two or more lists will be biased in assignment toward the most recent list. The distribution of element types and the assignment biases shown in Figure 1 lead to the prediction that list 1 items will be lower in correct list identification than list 2 items, which in turn will be lower than list 3 items. Misidentifications of list 1 items should be assignments to lists 1 + 2 or list 2. Misidentifications of list 2 items should be assignments to lists 2 + 3 or list 3. Finally, misidentified list 3 items should be assigned to lists 2 + 3. A control condition with nonoverlapping lists should show higher accuracy of identification for all three item types and a different pattern of misidentifications.

Predictions can also be derived from Figure 1 for items appearing in more than one list. The items in lists 1 + 2 should be lower in correct list identification than those in lists 2 + 3 and lists 1 + 2 + 3, which may



Item Type	Element Types	Assignment Bias		
		List 1	List 2	List 3
List 1	a + ab + abc	a	ab	abc
List 2	b + ab + bc + abc		b, ab	bc, abc
List 3	c + bc + abc			c, bc, abc
Lists 1 + 2	a + b + 2ab + bc + 2abc	a	b, ab	bc, abc
Lists 2 + 3	b + c + ab + 2bc + 2abc		b, ab	c, bc, abc
Lists 1+ 2 + 3	a + b + c + 2ab + 2bc + 3abc	a	b, ab	c, bc, abc

Figure 1. A graphic representation of the relationships among pools of contextual elements and specific element types and their association with different item types

not differ. Misidentified items in lists 1 + 2 should be assigned to lists 1 + 2 + 3 or 2 + 3. Misidentified items in lists 2 + 3 should be assigned to list 3 or lists 1 + 2 + 3. Finally, misidentified items in lists 1 + 2 + 3 should be assigned to lists 2 + 3.

METHOD

Design

The present design involved three groups, two experimental and one control, that learned three successive lists for free recall. In the *experimental* conditions, the variable manipulated was knowledge of the interlist relationships, with subjects either *informed or uninformed* about the degree of overlapping of items across successive lists. Each of the three successive lists for free recall consisted of 32 items, and the difference between the experimental conditions and the control condition was a function of the lists' overlap. In the *control* condition, the three successive lists contained *no* overlapping items. In the two experimental conditions, half of the list 1 items were included in list 2 together with an equal number of new items. List 3 was composed of half of the list 2 items together with an equal number of new items. The list 2 items included in list 3 were divided between items that had only appeared in list 2 and items that had appeared in both lists 1 and 2. Thus, in the experimental conditions, there was an additional within-subjects factor of item type for lists 2 and 3. In list 2, there were both old and new items, and in list 3, there were new, old (from list 2), and old (from lists 1 + 2) items.

Materials

The items used to construct the lists for free recall consisted of 96 single-syllable nouns divided into 12 subsets of 8 items each, balanced for frequency and imagery. In the construction of the 32-item lists, 12 possible combinations of the 8-item subsets were used. Each subset of 8 items and each of the 12 combinations of subsets was used equally often for list 1, list 2, and list 3 learning both within and between conditions.

Procedure

Upon entering the experimental situation, the subjects were told that they would see a list of words individually presented at a 2-sec rate by a projector and that their task was to learn the items so that at the end of presentation they could recall them within a 2½-min period of written recall. The subjects were further told that they could recall the words in any order they wished and that the study/test procedure would continue for several trials. After eight trials on list 1, the subjects were told that they would then see a new list that would be tested in the same way as the preceding list. The subjects in the informed condition were also told of the exact relationship between lists 1 and 2. After eight trials on list 2, the subjects were then told that they would see and be tested on another new list. The informed subjects were again told the exact relationship between list 3 and lists 1 and 2. List 3 learning continued for eight trials.

After list 3 learning, the subjects were given a booklet in which all 96 nouns from the total pool were listed in random order. Next to each word were the

symbols 1, 2, 3, 1 + 2, 2 + 3, 1 + 2 + 3, and N. The subjects were told to decide for each word listed whether they had seen that word in the first list only (1), the second list only (2), the third list only (3), the first and second lists (1 + 2), the second and third lists (2 + 3), all three lists (1 + 2 + 3), or never (N). The subjects were allowed unlimited time for this test of list identification.

Subjects

The subjects were 72 University of Pittsburgh undergraduates participating to fulfill a class requirement. The subjects were assigned to conditions using a block-randomization procedure to ensure that within each block of three subjects each condition appeared once.

RESULTS AND DISCUSSION

LIST 1

The analysis of percent correct recall on each trial of list 1 learning showed no effect of conditions [$F < 1$] and no interaction of conditions by trials. Performance on the last trial was not differential across conditions, with an overall mean of 76% correct. Thus, the three groups were equivalent in performance before list 2 learning.

LIST 2

Overall recall

The experimental and control groups were compared in an analysis of percent correct recall averaged across all items. The results of this analysis can be seen in Figure 2. There was a significant interaction of conditions by trials [$F(14, 483) = 2.21, p < .01$], which reflects the change from initial positive or zero transfer to subsequent negative transfer in the experimental conditions. The overall effect of conditions was not significant [$F(2, 69) = 2.39, p > .05$]. However, a separate comparison showed that the experimental informed condition, 66%, was superior to the uninformed condition, 59% [$F(1, 69) = 4.59, p < .05$]. This analysis of overall list 2 recall supports the prediction that knowledge of the interlist relationship does not produce positive transfer with partially overlapping lists but instead reduces the level of negative transfer relative to an uninformed condition, similar to the results of Petrich et al. (1975).

Old and new items

List 2 recall of old and new items was analyzed in both experimental conditions. The results of this analysis can be seen in Figure 3. There was

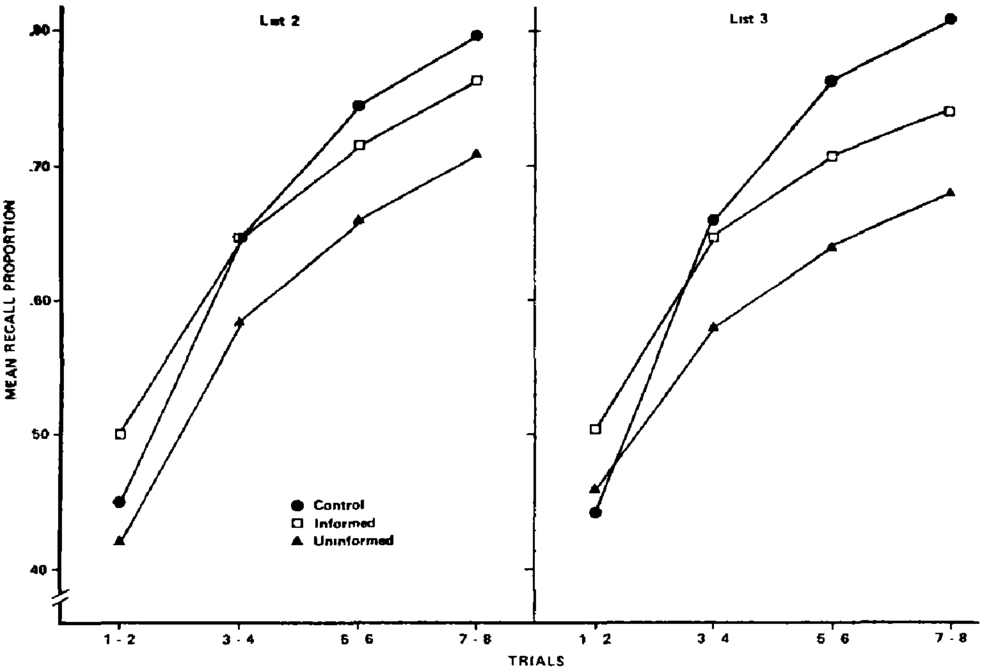


Figure 2. Mean recall proportion (percent correct) on list 2 and list 3 as a function of conditions and trials

a highly significant effect of item type [$F(1, 46) = 17.60, p < .001$], with new items, 66%, superior to old items, 60%. This superiority on the new items occurred in both the informed and uninformed conditions, with no interaction of conditions by item type [$F < 1$]. Thus, the analysis of item types is also consistent both with the predictions of Anderson and Bower's model and with the results obtained by Petrich et al. (1975).

LIST 3

Overall recall

An analysis of percent correct recall averaged across all items produced a pattern of results that can be seen in Figure 2. There was a highly significant interaction of conditions by trials [$F(14, 483) = 5.45, p < .001$], which again reflects the change from initial positive or zero transfer to subsequent negative transfer in the experimental conditions. The overall effect of conditions was again not significant [$F(2, 69) = 2.14, p > .05$], but a separate comparison again showed that the experimental informed condition, 66%, was marginally superior to the uninformed condition,

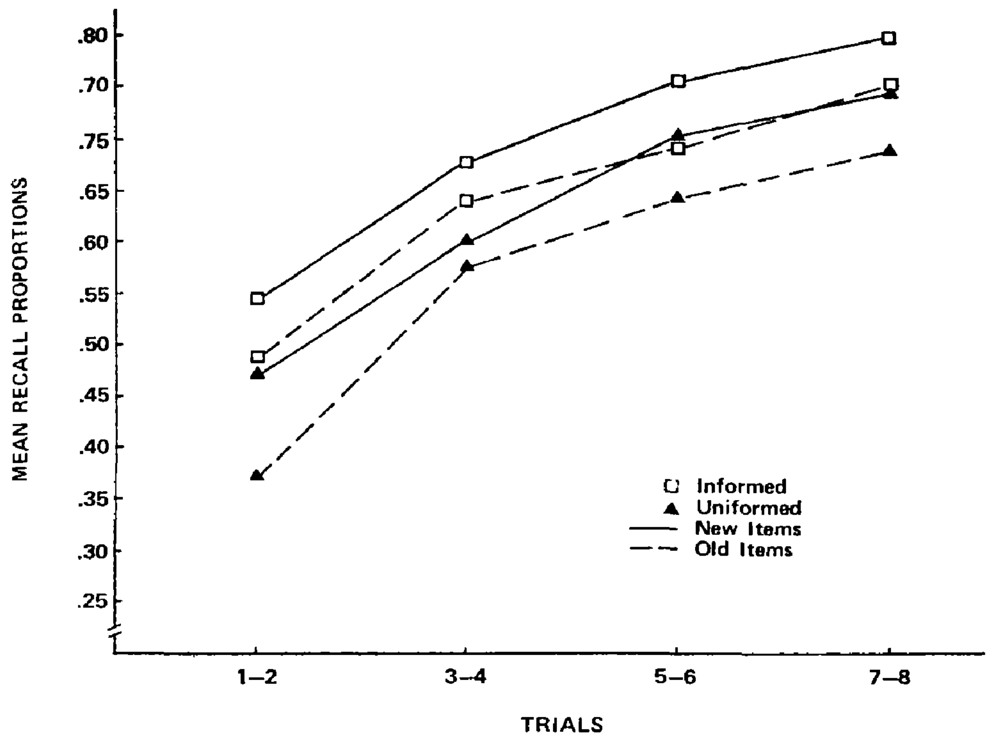


Figure 3. Mean recall proportion (percent correct) on list 2 as a function of experimental conditions, item type, and trials

59% [$F(1, 69) = 3.79, p = .06$]. This analysis is similar to the analysis of list 2 performance and again shows that knowledge of the interlist relationship does not produce positive transfer with partially overlapping lists but instead reduces the level of negative transfer relative to an uninformed condition.

Old and new items

The experimental conditions were again analyzed for the recall of old and new items. The results of this analysis can be seen in Figure 4. There was a significant effect of item type [$F(2, 92) = 3.72, p < .05$]. New items, 63.7%, did not differ from old items that appeared only in list 2, 62.6%; but both differed from old items that appeared in lists 1 + 2, 59.5%. Additionally, there was a marginally significant interaction of item type by trials [$F(14, 644) = 1.68, p = .06$], which can be seen in Figure 4. This interaction reflects the change in performance on old items

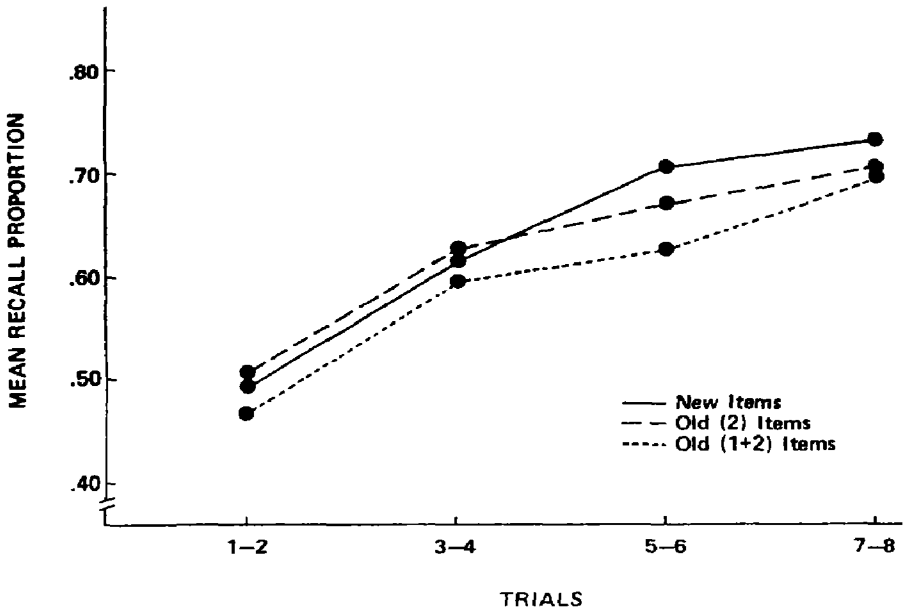


Figure 4. Mean recall proportion (percent correct) on list 3 as a function (type and trials

that appeared only in list 2 relative to new items, indicating a shift slight positive transfer to subsequent negative transfer during the s half of the list's acquisition [$p < .05$]. The effect of item type and its action with trials did not interact with experimental conditions [*F*s This analysis of item types supports the prediction that items app in multiple prior lists should show the poorest acquisition. The on tially discrepant result was the lack of a significant overall differen tween new items and old items that only appeared in list 2. Ho the interaction of item type by trials indicates that performance on old items was depressed relative to new items during the second h the list's acquisition.

LIST IDENTIFICATION

Performance on the test of list identification was scored for both c correct recognition and correct recognition plus appropriate list a ment. Overall correct recognition was high, ranging from 95% tc for all item types. The majority of conditions were above 98%, with est performance in the recognition of list 1 items (95%). Similar rate of false alarms for distractors was very low (1%). Table 1 pı

Table 1. Mean percent correct list identifications and errors as a function of item type and conditions

Item type	Correct (%)	Errors (distribution by lists, %)						
		1	2	3	1 + 2	2 + 3	1 + 2 + 3	N
Control								
List 1	67		*80*	5	6	3	1	5
List 2	81	*47*		*30*	4	8	4	7
List 3	89	7	*55*		12	13	5	8
Experimental								
List 1	33		*30*	2	*49*	8	4	8
List 2	50	8		*32*	15	*35*	8	2
List 3	79	1	12		5	*57*	21	4
Lists 1 + 2	56	7	16	4		*26*	*43*	4
Lists 2 + 3	63	0	11	*36*	7		*44*	2
Lists 1 + 2 + 3	67	1	3	6	16	*72*		3

the percent correct list assignments and the distributions of errors made on the various item types. An initial analysis compared performance on the list 1, list 2, and list 3 item types across the experimental and control conditions. This analysis produced a significant effect of conditions [$F(2, 69) = 10.86$, $p < .001$], with the control condition, 79%, superior to both the experimental informed, 55%, and uninformed, 54%, conditions. The latter two did not differ.[1] The effect of item type [$F(2, 138) = 82.80$, $p < .001$] showed list 3 items, 83%, superior to list 2 items, 60%, which were superior to list 1 items, 44%. Additionally, there was an interaction of item type by conditions [$F(4, 138) = 4.46$, $p < .01$], which can be seen in Table 1. The control condition was superior to the experimental conditions on all three item types, but with larger differences on the list 1 and list 2 item types. The two experimental conditions did not show differential performance across the three item types. Thus, as predicted, list identification was impaired in the experimental conditions, particularly on item types from the initial two lists.

The second analysis of list identification was restricted to the experimental conditions and included all six item types. This analysis showed that experimental conditions did not differ in overall list identification, nor did conditions interact with the item types [Fs < 1]. There was a highly significant effect of item type [$F(5, 230) = 24.84$, $p < .001$], which can be seen in Table 1. A subsequent Newman-Keuls test of individual item-type means showed that for items appearing in only a single list, performance on list 1 items was inferior to that on list 2 items, which was inferior to that on list 3 items. For those items appearing in multiple lists, performance on items in lists 1 + 2 was inferior to that on items in lists 2 + 3 and lists 1 + 2 + 3. The latter two did not differ. Additionally, items appearing only in list 3 were superior to all other item types. This pattern of results in the experimental conditions is also consistent with the predictions derived from Anderson and Bower's model (1972).

The final data supporting the predictions derived from that model involve the distributions of errors shown in Table 1. The experimental and control conditions clearly differ in their patterns of list assignments for items appearing in only a single list. In the control condition, the misassignment of item types seldom reflected a multilist assignment; in the experimental conditions, it did. The misassignment of item types in the experimental conditions is consistent with the predictions based on element types and their associated assignment biases.

To summarize, the present data on transfer and list identification, combined with those reported by Petrich et al. (1975), strongly suggest that the decision component is the major factor affecting the free recall of

successive overlapping lists. This decision component is best described by Anderson and Bower's model (1972) of the roles of list tagging and contextual elements in that recall. Such a model appears applicable to transfer on part/whole, whole/part, and partially overlapping lists, as well as to list identification of items on multiple lists (Petrich et al., 1975; Sternberg and Bower, 1974).

Notes

This research was supported by funds provided to the first author by the Learning Research and Development Center, University of Pittsburgh, which is funded in part by the National Institute of Education (NIE). The authors thank Penny Parkinson and Linda Swanson, who greatly assisted in the data collection and tabulation. Requests for offprints should be addressed to James W. Pellegrino, Learning Research and Development Center, University of Pittsburgh, Pittsburgh, Pennsylvania 15260.

1. The lack of a difference in list identification between the two experimental groups is not inconsistent with the difference in recall nor with the data obtained by Petrich et al. (1975). Both experimental groups showed parallel effects for recall of old and new items, and differed only in absolute levels of recall. The lack of a similar difference in list identification indicates that additional factors may be contributing to the difference in recall (e.g., organizational strategies or differences in response criteria, as suggested by Petrich et al., 1975).

References

Anderson, J. R., and Bower, G. H. 1972. Recognition and retrieval processes in free recall. *Psychological Review* 79:97–123.

Bower, G. H., and Lesgold, A. M. 1969. Organization as a determinant of part-to-whole transfer in free recall. *Journal of Verbal Learning and Verbal Behavior* 8:501–506.

Novinski, L. 1972. A reexamination of the part-whole effect in free recall. *Journal of Verbal Learning and Verbal Behavior* 11:228–233.

Petrich, J. A., Pellegrino, J. W., and Dhawan, M. 1975. The role of list information in free-recall transfer. *Journal of Experimental Psychology: Human Learning and Memory* 1:326–336.

Schulze, N., and Gorfein, D. S. 1976. Subject strategies in part-whole transfer in free recall. *Memory and Cognition* 4:311–317.

Sternberg, R. J., and Bower, G. H. 1974. Transfer in part-whole and whole-part free recall: A comparative evaluation of theories. *Journal of Verbal Learning and Verbal Behavior* 13:1–26.

AMERICAN JOURNAL OF PSYCHOLOGY
September 1977, Vol. 90, No. 3, pp. 431-446

Concurrent processing demands and the experience of time-in-passing

Robert E. Hicks, George W. Miller, Gerald Gaes, and Karen Bierman
State University of New York at Albany

Under the prospective paradigm, judged time decreased monotonically with the increased processing demands of concurrent card sorting (Experiment I) and of concurrent verbal rehearsal (Experiment II). It was nonmonotonically related to concurrent tapping rate (Experiment III), which latter, when required during verbal rehearsal, had an identical curvilinear effect on short-term recall (Experiment IV). It is concluded that the experience of time-in-passing is an inverse function of the processing demanded by a concurrent task. An attentional model is suggested and evaluated against the literature.

In a recent paper, Hicks, Miller, and Kinsbourne (1976) reviewed the literature relating judged time to concurrent stimulation and information processing. Concerning *prospective* judgments of duration, those made when the subject knows in advance that the duration of an interval is to be judged, they reached two general conclusions. The first was that judged time increases directly with the number of stimuli and, less consistently, with the complexity of stimuli presented during the intervals to be judged; this conclusion holds only when no processing of the stimuli is required of the subject. The second conclusion was that when the subject is required to process the stimuli presented or perform some concurrent nontemporal task, judged time decreases. Both general effects are more pronounced with longer intervals.

The first general conclusion is supported by several parametric studies, but the second is based largely on studies that varied concurrent activity in a nonsystematic fashion. Some examples: Yerkes and Urban (1906) found that subjects estimated intervals spent listening or reading material as longer in duration than the same amount of time spent taking dictation. Similarly, Swift and McGeoch (1925) reported that intervals spent listening to a text were judged as longer in duration than the same intervals spent copying the text. Axel (1924) had subjects estimate the duration of

intervals of from 15 to 30 sec that were spent doing nothing, unpaced tapping, crossing out signs, finding analogies, and completing figures. The mean constant errors he reported were +1.8, +2.4, −5.7, −7.6, and −9.2 respectively. Gulliksen (1927) reported the following mean estimated durations of a 200-sec interval spent: resting, 241.7 sec; holding arms out, 228.4 sec; listening to a metronome, 218.9 sec; pressing a point on the skin, 210.2 sec; reading a passage, 181.8 sec; taking dictation, 174.6 sec; and performing division, 168.9 sec. Hicks and Brundige (1974) had subjects judge time while doing nothing else or while being simultaneously tested on their recognition memory of words or of faces. An extreme is provided by Loehlin (1959), who compared the judged time of intervals spent at 16 qualitatively different tasks.

In the present study, processing demands were systematically varied for each of several concurrent tasks. In all the experiments, subjects were told in advance that an estimate of time would be required (i.e., they made prospective judgments). Subjects were instructed not to count or otherwise "mark time" during the intervals but to translate their subjective experience of duration into seconds at the conclusion of each interval.

EXPERIMENT I

The concurrent task in this experiment was sorting playing cards into various classes. Crossman (1953) found that sorting time was linearly related to $\log_2$ of the number of sorting classes. Murdock (1965) had observers sort cards into one, two, or four classes during the (auditory) presentation of words for later recall. Recall of the words decreased monotonically with the number of sorting classes; that is, the relative processing demanded by the sorting can be inferred to increase with number of sorting classes.

METHOD

Subjects

The subjects were enrolled in introductory psychology courses at the State University of New York at Albany, and they participated in experiments as part of a course requirement. The 16 men and 16 women who participated in this experiment had not been in previous experiments on judged time.

Procedure

Subjects were informed that a series of intervals would be defined by the experimenter's saying "Start" and "Stop" and that they were to estimate in seconds the duration of the interval between the two commands. They were told

that they would also be sorting cards during some trials: sometimes (face up) into a single stack, sometimes into two stacks on the basis of color, and sometimes into four stacks on the basis of suit. They then demonstrated their understanding of this concurrent task by sorting several cards under each condition. Subjects removed their watches at the start of the experiment. Each subject was tested individually in a soundproof chamber 3 by 4 m, illuminated by four 20-W fluorescent tubes directly overhead.

Subjects judged time under four conditions. The order of conditions was controlled by a latin square with four subjects of each sex per order. Within each condition, three intervals (8, 13, and 22 sec) were administered in a separate block randomization for each subject and each condition. Thus, each subject had 12 judgment trials. Intervals were timed by stopwatch. In a block of card-sorting trials, subjects were handed a different shuffled deck of 52 playing cards (jokers removed) on each trial. At the end of each trial, the experimenter recorded the subject's judged time of the interval. On sorting trials, the experimenter also removed the stack(s) of cards for later counting and scoring.

Design

The design was a split plot with the sex of the subjects and the order of the conditions as between-subject variables and with *condition* (a concurrent card-sorting task with a response uncertainty of 0, 1, or 2 bits per response, plus a control with no concurrent task) and *interval* (8, 13, and 22 sec) as within-subject variables. The two within-subject variables were orthogonally decomposed, using pooled within-cell deviations from fit as error terms (see Grant, 1956). Coefficients for the unequal time intervals were generated using Gaito's (1965) procedure.

RESULTS

Card sorting

Subjects sorted means of 35.51, 28.69, and 19.03 cards in the 0-, 1-, and 2-bit conditions [$F(2, 48) = 172.166, p < .01$], and this effect was linear [$F(1, 24) = 523.40, p < .01$]. Subjects thus sorted at rates of 2.27, 2.00, and 1.33 cards per sec in the 0-, 1-, and 2-bit conditions. This yields 2 bits and 2.66 bits of information transmitted per second in the 1- and 2-bit conditions respectively.

Time judgment

Condition was a significant source of variation [$F(3, 72) = 4.752, p < .01$]. The mean judged times were 20.91, 19.54, 16.49, and 16.39 under the control, 0-, 1-, and 2-bit conditions respectively. The judged times under the control condition were longer than those under the combined card-sorting conditions [$F(1, 24) = 7.901, p < .01$], and judged times under the 0-, 1-, and 2-bit conditions decreased linearly with increased response uncertainty [$F(1, 24) = 7.861, p < .01$].

The effect of interval was significant [$F(2, 48) = 175.844, p < .01$], and the linear component accounted for this effect [$F(1, 24) = 585.813, p < .01$]. The mean judged times were 10.39, 16.93, and 27.66 for the 8-, 13-, and 22-sec intervals.

The linear component of interval differed for the control and combined card-sorting conditions: (control versus combined card-sorting conditions) $\times$ ($\text{interval}_{\text{linear}}$) [$F(1, 24) = 8.33, p < .01$]. And the slope of the function relating judged time to actual clock time decreased with increased response uncertainty in the card-sorting conditions: ($\text{response uncertainty}_{\text{linear}}$) $\times$ ($\text{interval}_{\text{linear}}$) component [$F(1, 24) = 6.79, p < .025$]. This interaction is presented in Figure 1. The slopes relating judged time to actual clock time were 1.3, 1.3, 1.2, and 1.1 in the control, 0-bit, 1-bit, and 2-bit conditions. No other component in the analysis produced a F significant at the .1 level.

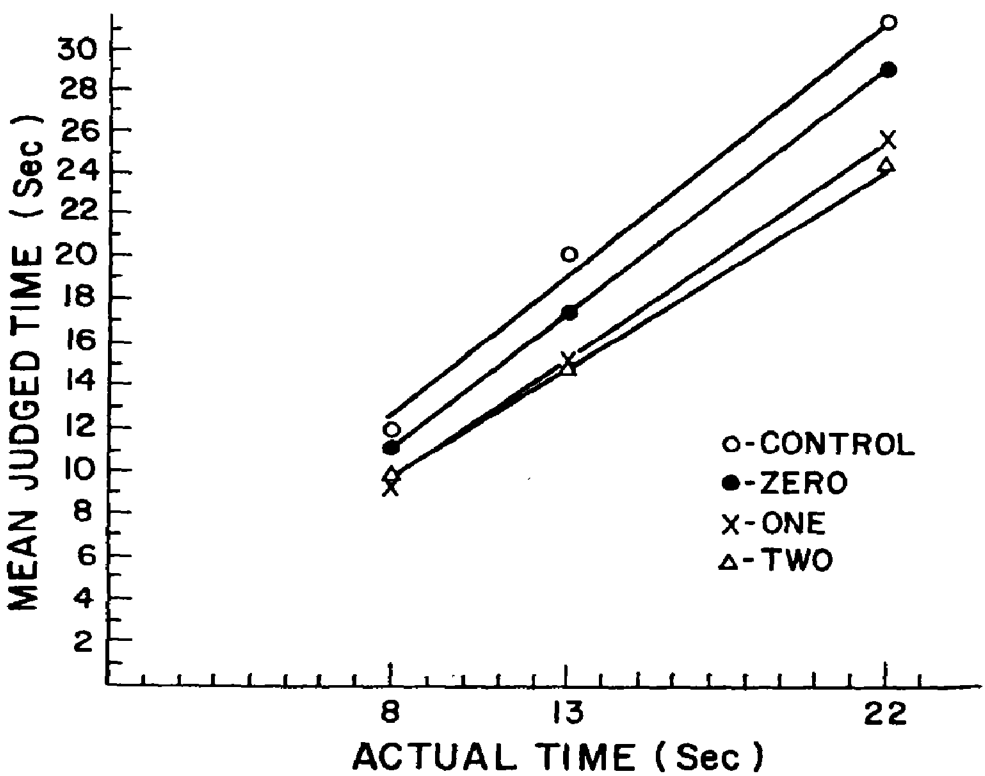


Figure 1. Judged time as a function of actual time and concurrent response uncertainty; Experiment I

DISCUSSION

Judged time decreased linearly with the increased processing demands of the concurrent task. This result supplements previous reports that concurrent activity decreases judged time. Hicks et al. (1976) used the same concurrent task in a between-subject design with a 45-sec interval. For the subjects making prospective judgments in that study too, judged time decreased linearly with increased response uncertainty. It is noteworthy that the same result was obtained in the two studies using different subjects, different designs, and different intervals. The effect was more pronounced in the study by Hicks et al. (1976), −10.958 (bit), than in the present study, −1.58 (bit).

Hicks et al. (1976) did not have a control condition with no concurrent task. The inclusion of that condition in the present study strengthens their interpretation that judged time decreases with the processing demanded by the concurrent activity.

EXPERIMENT II

The concurrent task in this experiment was verbal rehearsal. Rehearsal generally requires processing (Kerr, 1973), and the processing demanded has been found to increase with the difficulty of the items (Hicks, Provenzano, and Rybstein, 1975; Stanners, Meunier, and Headly, 1969). Hicks et al. (1975) found that sequential finger movements decreased in speed and accuracy when participants concurrently rehearsed eight-letter strings and that this decrement varied inversely with the redundancy of the strings of letters. In the present experiment, the rehearsal conditions used by Hicks et al. (1975) were embedded within the time intervals to be judged.

METHOD

Subjects

There were 15 men and 15 women, from the same population as in Experiment I, who served in this experiment. None of the subjects had previously participated in experiments on judged time.

Materials

All 15 of the eight-letter strings at each of the zero, first, second, and fourth orders of approximation to English (Miller, Bruner, and Postman, 1954) were used. These strings have 0%, 15%, 29%, and 43% redundancy respectively. They were presented on a memory drum.

Procedure

Each subject was tested individually, in the same room used for Experiment I. The subject was seated in front of a table on which the memory drum was placed. The subject was told that he/she should look at the window of the memory drum immediately following a 'ready' signal on each trial. If a row of asterisks appeared, the only task was to judge time. If a string of letters appeared, the subject was instructed to memorize them for subsequent serial recall.

The asterisks or letters always remained exposed for 4 sec, followed by a blank. The interval to be judged began when the blank field was exposed and ended when the experimenter said "Stop." At the end of the interval, the subject orally reported an estimate, in seconds, of the interval's duration and, on rehearsal trials, attempted written recall of the letters. After the instructions, each subject had removed his/her watch and practiced twice with 43%-redundant strings of letters.

Each subject had 15 trials: rehearsal with each of the four levels of redundancy, and a no-rehearsal control, at each of three intervals (8, 16, and 32 sec). The order of the 15 trials was completely randomized independently for each of the 15 subjects within a sex. Across sexes, the same randomization of orders was used. The intervals were measured by stopwatch. On rehearsal trials, subjects were instructed to rehearse the letters during the retention intervals. Different subjects received different exemplars of letters at each level of redundancy.

Design

Sex was the only between-subject variable. *Interval* (8, 16, and 32 sec) and *condition* (rehearsal with 43%-, 29%-, 15%-, and 0%-redundant strings of letters, plus a control with no rehearsal) were within-subject variables. Effects were orthogonally decomposed as previously described.

RESULTS

List recall

Redundancy was the only significant effect in this analysis [$F(3, 28) = 4.32, p < .01$], and only the linear component of this effect [$F(1, 28) = 9.673, p < .01$]. The mean numbers of errors (omissions and sequencing errors) were 2.4, 1.4, .6, and .3 for the 0%-, 15%-, 29%-, and 43%-redundant conditions respectively.

Time judgment

The difference between intervals was significant [$F(2, 56) = 322.836, p < .01$], and the linear component accounted for the effect [$F(1, 28) = 356.49, p < .01$]. Mean judged times were 8.61, 15.99, and 28.79 for the 8-, 16-, and 32-sec intervals.

The difference between conditions was not significant at an acceptable level [$F(4, 112) = 2.239, p < .1$]. Neither the comparison of control versus combined rehearsal conditions [$F(1, 28) = 2.907$] nor the linear

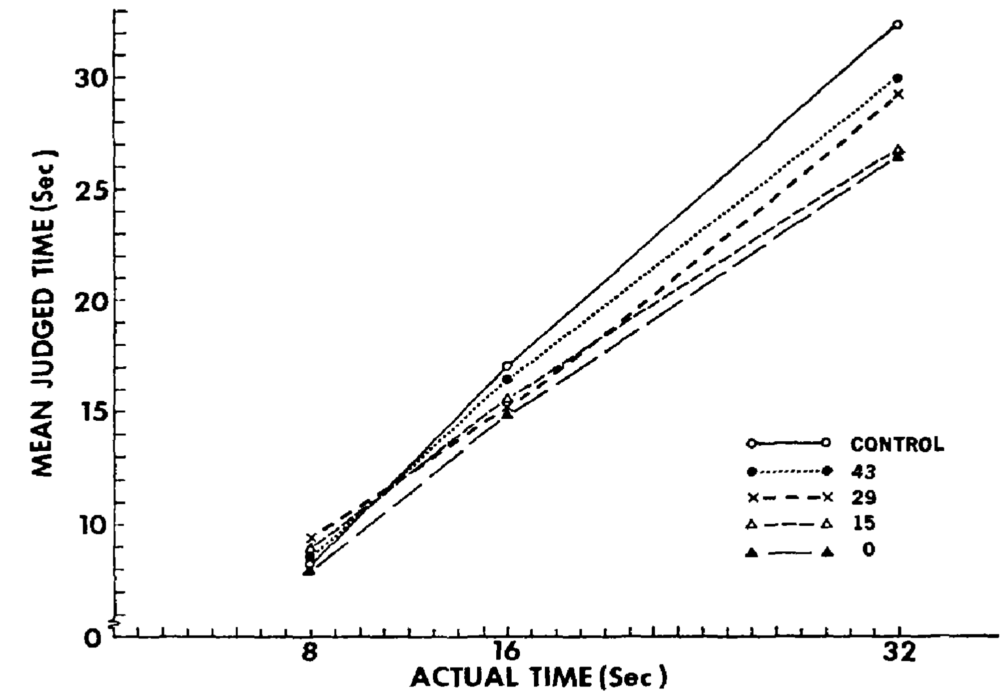


Figure 2. Judged time as a function of actual time and concurrent rehearsal condition; Experiment II

component of the rehearsal conditions [$F(1, 28) = 3.392$] reached an acceptable level of significance, but the combination of these two components was significant [$F(2, 56) = 3.15, p \simeq .05$]. These two components accounted for 98.38% of the sum of squares for condition. The mean judged times were 19.17 in the control condition, and 18.26, 17.98, 17.06, and 16.52 for the 43%- through 0%-redundancy conditions, in order.

Interval and condition produced a significant interaction [$F(8, 224) = 3.00, p < .01$]. This interaction is presented in Figure 2. Approximately 99% of the sum of squares of this interaction is accounted for by two components: (control versus combined rehearsal conditions) × ($\text{interval}_{\text{linear}}$) [$F(1, 28) = 7.925, p < .01$] and, in the rehearsal conditions, ($\text{redundancy}_{\text{linear}}$) × ($\text{interval}_{\text{linear}}$) [$F(1, 28) = 4.216, p < .05$]. The slopes of the functions relating judged time to actual clock time were .995 for the control condition and .872, .84, .732, and .752 for the 43%- through 0%-redundancy conditions, in order.

DISCUSSION

Judged time decreased monotonically with the decreased redundancy (increased processing demands) of the concurrently rehearsed lists. This result is consistent with the hypothesis that the judgment of time requires processing capacity. Hicks et al. (1975) found that motor performance decreased monotonically with decreased redundancy of the (same) lists concurrently rehearsed, so an independent assessment of processing demand is available for the rehearsal task. Hicks and Brundige (1974) similarly found that judged time decreased during a test of recognition memory, as compared to a control with no such test, and that this effect increased with longer intervals.

EXPERIMENT III

This experiment varied the rate at which subjects were required to tap during the presentation of the intervals whose durations they were to judge. Considering the relationship between movement time, movement distance, and target width (Fitts, 1954), it was predicted that judged time would decrease monotonically as concurrent tapping rate increased.

METHOD

Subjects

There were 16 men and 16 women, from the same population as in the previous experiments, who served as subjects. None of them had served in other studies of judged time.

Procedure

Each subject was tested individually, in the same room used for the previous experiments. The subject was seated in front of a table on which an electric metronome, a tapping board, and an event counter were placed. The tapping board was 20 by 6 cm, with two 6-cm-square metal plates at each end. A metal stylus (with rubber handgrip) was connected with the metal plates and the event counter such that one contact of stylus and plate incremented the counter one unit. The subject was told that he/she would be making a series of judgments of time and that concurrent paced tapping would be required on some trials. The subject then practiced tapping at 48, 96, and 144 beats per min, attempting to make the contacts coincide with the metronome clicks.

The four conditions (tapping at 0, 48, 96, or 144 beats per min) were presented in blocks according to a 4 × 4 latin square with four subjects of each sex assigned to each order. Within each block, 16 independent randomizations of 12-, 19-, 32-, and 54-sec intervals were prepared. Each subject thus had 16 judgment trials. One male and one female subject were assigned to each order of intervals.

Design

The sex of the subjects and the order of conditions were between-subject variables, and *condition* (concurrent tapping at 48, 96, or 144 beats per min, plus the 0-beat control with no tapping) and *interval* (12, 19, 32, and 54 sec) were within-subject variables. Effects were orthogonally decomposed as previously described.

RESULTS

Tapping

The mean interval presented was 29.25 sec. The numbers of taps expected would thus be 23.4, 46.8, and 70.2 for the 48-, 96-, and 144-beat conditions respectively. The mean numbers of taps recorded were 23.23, 43.89, and 66.46 for those three conditions [$F(2, 48) = 14{,}825.19$, $p < .01$; $F_{linear}(1, 24) = 37{,}468.95$, $p < .01$]. Thus, the subjects tapped at approximately the required rates.

Time judgment

The effect of interval was significant [$F(3, 72) = 298.661$, $p < .01$], and the linear component accounted for 99.83% of the sum of squares [$F(1, 24) = 1{,}231.224$, $p < .01$]. The mean judged times were 12.04, 18.12, 29.66, and 52.32 for the intervals of 12, 19, 32, and 54 sec, in that order.

Condition affected judged time [$F(3,72) = 4.825$, $p < .01$]. This effect is presented in the top curve of Figure 3. Only the quadratic component of this effect reached significance [$F(1, 24) = 23.328$, $p < .01$]. By least squares, judged time = 27.65 − .04 (tapping rate) + .0004 (the square of tapping rate). None of the other effects reached significance at the .1 level.

DISCUSSION

Judged time was found to be a quadratic function of concurrent responding in this experiment. Parker (1973) reported a similar result: judged time (for a 2-sec interval) was a curvilinear function of concurrent induced muscle tension, with relative underestimation of actual clock time occurring only at 50% of maximum tension, and not at 25% or 75%.

In order to interpret the results of this experiment and those of Parker (1973), it is necessary to obtain an independent assessment of the processing demands of the concurrent tasks. The relative processing demands of the card-sorting tasks employed in Experiment I were defined by Murdock (1965) using concurrent verbal memorization. Hicks et al. (1975)

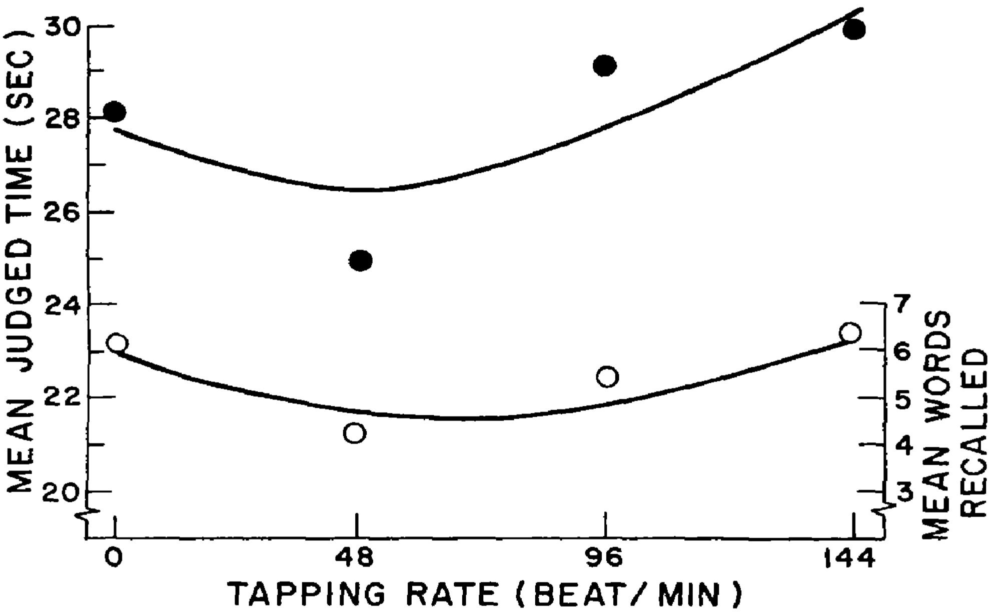


Figure 3. The top curve/left ordinate represents the effect of concurrent tapping rate on judged time; Experiment III. The bottom curve/right ordinate presents the number of words recalled as a function of tapping rate during rehearsal; Experiment IV

demonstrated the relative processing demands of the rehearsal conditions of Experiment II using concurrent motor performance.

The relative processing demands of the present tapping tasks, as well as the ergographic tasks of Parker (1973), are not established independent of judged time. Experiment IV therefore, was designed to measure the interference of concurrent tapping tasks with verbal rehearsal.

EXPERIMENT IV

METHOD

Subjects

The subjects were 8 men, from the same population as in the previous experiments.

Materials and procedure

Four lists of 12 nouns each were constructed. All words had frequency counts of less than 10 per million (Thorndike and Lorge, 1944). The lists were presented on a memory drum.

Each subject was individually tested, in the same room used for the previous

experiments. The subject was seated at a table on which was the memory drum. He was told that he would receive four trials, each composed of a list's presentation, a retention interval, and a period for (written) free recall of the words. He was told to study the words as they were presented in the window of the memory drum (1-sec rate) and to rehearse the words during the retention interval (25 sec).

He was told that on some trials he would be required to tap during the retention interval. The three rates of tapping were demonstrated for him, and he practiced each one for approximately 25 sec. For any given tapping trial, the subject was presented with the rate he would be tapping prior to the start of the trial.

The subject then received the four trials. Each subject had each list presented once, and performed once under each condition (tapping at 0, 48, 96, or 144 beats per min). The order of lists and conditions was controlled via a latin-greco square such that across subjects each list occurred equally often with each condition. There were two subjects assigned to each sequence.

The design was thus a split plot with *order* of conditions as a between-subject variable and *conditions* per se (concurrent tapping at 48, 96, or 144 beats per min, plus a 0-beat control with no tapping) as a within-subject variable.

RESULTS

Tapping

The mean numbers of taps per trial were 21.5, 40.375, and 61.375 for the 48-, 96-, and 144-beat conditions respectively [$F(3, 8) = 1{,}212.627$, $p < .01$]. This effect was linear [$F(1, 4) = 5{,}834.917$, $p < .01$], and no other F was larger than 1.0. Thus, subjects performed as expected. The numbers of taps expected would be 20, 40, and 60 for a 25-sec interval.

Recall

The number of words recalled was scored for each subject on each trial. The only significant effect on amount recalled was condition [$F(3, 12) = 7.20$, $p < .01$]. This effect is presented in the lower curve of Figure 3. Orthogonal decomposition of this effect yielded only a significant quadratic component [$F(1, 4) = 26.889$, $p < .01$]. By least squares, number recalled = 5.96 − .04 (tapping rate) + .003 (the square of tapping rate). The constants of −.04 and +.0003 are remarkably similar to those relating judged time to tapping rate, −.04 and +.0004. No other effect or component was significant at the .1 level.

DISCUSSION

The effect of concurrent tapping rate was virtually identical for judged time (Experiment III) and words recalled (Experiment IV). That is to

say, the processing demands of the tapping tasks seem identical when paired with the estimation of temporal duration or with the rehearsal of lists of words.

GENERAL DISCUSSION

Verbal estimation of temporal duration (from subjects set to judge time) decreases monotonically with the processing demands of concurrent tasks. The present results are consonant with this generalization, which brings some order to the numerous reports of the differing estimations of the temporal durations of widely differing activities. This hypothesis was implied by several psychologists and philosophers over the years: James (1890), Lavelle (1945), Romanes (1878), and Wundt (1874) all suggested that the experience of duration increases with the amount of attention devoted to time, as such, during the interval. It follows that distraction from time by the concurrent processing of nontemporal information should decrease the experience of duration.

This basic thesis has been very elegantly expressed by several authors. McTaggert (1927, p. 277), for example, claimed that when "we pay as much attention to time in a short period as we should usually pay in a longer period, we judge the period to be longer." While absorbed in some activity, conversely, "we have little attention to spare for the lapse of time, and so we judge that little time has elapsed." According to Sturt (1925, p. 89), "the estimates of duration seem to depend on a certain division of attention, [for] we both experience and reflect on the succession of experiences; and if this division of attention is destroyed, as when we become excitedly immersed in some occupation, we are apt to become oblivious of the passage of time."

Virtually all more formal theories of the 'time sense' have, explicitly or implicitly, assumed the following processes (this description is for verbal estimation, but the basic ideas would not change for the other methods): a *time base,* which is responsible for subjective temporal units; a *counter,* which stores these units until the conclusion of the interval; and a *response translator,* which maps the contents of the counter onto the response surface (usually, conventional time units). Different theorists have assumed the time base is predicated upon metabolism, heart rate, respiration cycle, alpha rhythm, cerebral processes, cerebellar processes, the sensorimotor feedback cycle, or neural scanning, among a great many other possibilities (Michon, 1967, has an excellent review).

The time base hypothesized in this paper comes from Frankenhaeuser (1959), who considers the subjective unit of time to be based on the

overlearning of an "average mental content per unit of duration." The number of mental events, stimulus- or self-generated, experienced during an interval is the basis of the observer's judgment. The present position is identical to Frankenhaeuser's except for our distinction between events that can be integrated with the time-judgment task (e.g., the mere presentation of stimuli) and events orthogonal to the time-judgment task (e.g., a concurrent nontemporal task). This distinction rationalizes the opposite effects of information presented and information processed as reviewed by Hicks et al. (1976). Thus, increasing the number of events adds more into the counter, causing an increase in judged time, whereas information processing prevents the storage of events into the counter, causing a decrease in judged time. The complexity of the input is not addressed by this hypothesis. Stimulus complexity has a far from consistent effect on judged time, however. It may well depend on how the subject 'chunks' the input into subjective events; that is, the nominal and functional stimuli may differ somewhat.

There is evidence for the learning of temporal units. Orsini (described by Fraisse, 1963) had seven-year-olds attempt to reproduce 30-sec intervals. Initially, 9.6% of them were correct to within 5 sec, 78% of them had reproductions longer than 35 sec, and 12.4% of them had reproductions shorter than 25 sec. The children were then given three weeks of training (with informative feedback about their errors) in judging time. The training, of course, decreased the errors. But, more importantly, tests (without the informative feedback) that were given three months after the training showed that the children were much like untrained adults in their reproductions: 40.9%, 29.1%, and 30% of the children produced correct (to within 5 sec), longer, and shorter reproductions respectively. These figures were 36%, 29%, and 35% for untrained adults. The training had apparently speeded up the acquisition of temporal units for the children, providing them with temporal representations usually informally acquired later in development.

To summarize, the experience of duration (under the paradigm using prospective judgments) increases with the observer's *attention* to time. This is because the events on which the experience of time-in-passing is based require attention (processing capacity) for storage. Stimuli requiring no processing can increase experienced duration by increasing the number of events in storage. Stimuli requiring processing can decrease experienced duration because fewer of the events defining duration are stored.

This attentional approach is consonant with several other effects often reported. A large number of studies report that without informative feed-

back, experienced time decreases over trials (see Hicks et al., 1976). Several authors have pointed out the similarity of this effect to the frequently reported vigilance decrement over trials (e.g., Hawkes and Sherman, 1972; von Sturmer, 1966, 1968), a decrement usually attributed to a decrease in focal attention (Mackworth, 1969). For example, amphetamine and other stimulants, which have been hypothesized to increase focal attention (Talland and Quarton, 1966; Weiner and Ross, 1962a), both increase judged time (Frankenhaeuser, 1959; Weiner and Ross, 1962b), and reduce the vigilance decrement (Mackworth, 1969). Sedatives such as barbiturates and alcohol, which have been hypothesized to diffuse attention (Mirsky and Kornetsky, 1964), cause an increase in the vigilance decrement (Mackworth, 1969) and a decrease in judged time (Frankenhaeuser, 1959). It has also been found that extraverts show poorer performance than introverts on tasks involving sustained attention (Bakan, Belton, and Toth, 1963; Broadbent, 1958; Corcoran, 1965; Davies and Hockey, 1966); they also judge time as shorter than introverts (Claridge, 1967; Eysenck, 1959; Lynn, 1967). The attentional hypothesis appears consistent with most reported effects in prospective judgments of time.

Notes

The authors thank Patricia L. Hicks for typing this manuscript. Requests for offprints should be sent to the first author, Psychology Department, State University of New York at Albany, Albany, New York 12222. Revision received for publication March 18, 1976.

References

Axel, R. 1924. Estimation of time. *Archives of Psychology* 12, no. 74.

Bakan, P., Belton, J. A., and Toth, J. C. 1963. Extraversion-introversion and decrement in an auditory vigilance task. In *Vigilance: A Symposium*, ed. D. N. Buckner and J. J. McGrath. New York: McGraw-Hill.

Broadbent, D. A. 1958. *Perception and communication*. New York: Pergamon Press.

Claridge, G. S. 1960. The excitation-inhibition balance. In *Experiments in personality*, vol. 2, ed. H. J. Eysenck. London: Routledge and Kegan Paul.

Corcoran, D. W. J. 1965. Personality and the inverted-U relation. *British Journal of Psychology* 56:267–273.

Crossman, E. R. F. W. 1953. Entropy and choice time: The effect of frequency unbalance on choice-response. *Quarterly Journal of Experimental Psychology* 5:41–51.

Davies, D. R., and Hockey, G. R. J. 1966. The effects of noise and doubling the signal frequency on individual differences in visual vigilance performance. *British Journal of Psychology* 57:381–389.

Eysenck, H. J. 1959. Personality and the estimation of time. *Perceptual and Motor Skills* 9:405–406.

Fitts, P. M. 1954. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology* 47:381–391.

Fraisse, P. 1963. *The psychology of time.* New York: Harper & Row.

Frankenhaeuser, M. 1959. *Estimation of time: An experimental study.* Stockholm: Almqvist & Wiksell.

Gaito, J. 1965. Unequal intervals and unequal *n* in trend analyses. *Psychological Bulletin* 63:125–127.

Grant, D. A. 1956. Analysis-of-variance tests in the analysis and comparison of curves. *Psychological Bulletin* 53:141–154.

Gulliksen, H. 1927. The influence of occupation upon the perception of time. *Journal of Experimental Psychology* 10:52–59.

Hawkes, G., and Sherman, S. 1972. Vigilance effects for duration judgments with two levels of task demand. *Perceptual and Motor Skills* 34:351–356.

Hicks, R. E., and Brundige, R. 1974. Judgments of temporal duration while processing verbal and physiognomic stimuli. *Acta Psychologica* 38:447–453.

Hicks, R. E., Miller, G. W., and Kinsbourne, M. 1976. Prospective and retrospective judgments of time as a function of amount of information processed. *American Journal of Psychology* 89:719–730.

Hicks, R. E., Provenzano, F. J., and Rybstein, E. D. 1975. Generalized and lateralized effects of concurrent verbal rehearsal upon performance of sequential movements of the fingers by the left and right hands. *Acta Psychologica* 39:119–130.

James, W. 1890. *The principles of psychology,* vol. 1. New York: Henry Holt and Co.

Kerr, B. 1973. Processing demands during mental operations. *Memory and Cognition* 1:401–412.

Lavelle, L. 1945. *Du temps et de l'eternite.* Paris: Aubier.

Loehlin, J. C. 1959. The influence of different activities on the apparent length of time. *Psychological Monographs: General and Applied* 73(4, whole no. 474).

Lynn, L. 1967. Clinical manifestations of psychiatric disorder. In *Comprehensive textbook of psychiatry,* ed. A. M. Freeman and H. I. Kaplan. Baltimore: Williams & Wilkins.

Mackworth, J. F. 1969. *Vigilance and habituation: A neuropsychological approach.* Middlesex, England: Penguin Books.

McTaggart, J. M. E. 1927. *The nature of existence,* vol. 2. Cambridge: University Press.

Michon, J. A. 1967. *Timing in temporal tracking.* Soesterberg, Netherlands: Institute for Perception.

Miller, G. A., Bruner, J. S., and Postman, L. 1954. Familiarity of letter sequences and tachistoscopic identification. *Journal of General Psychology* 50:129–139.

Mirsky, A. F., and Kornetsky, C. 1964. On the dissimilar effects of drugs on the symbol substitution and continuous performance tests. *Psychopharmacologia* 5:161–177.

Murdock, B. B. 1965. Effects of a subsidiary task on short-term memory. *British Journal of Psychology* 56:413–419.

Parker, N. K. 1973. Influence of induced muscular tension on a time-estimation motor task. *Journal of Motor Behavior* 5:111–120.

Romanes, G. J. 1878. Consciousness of time. *Mind* 3:179–203.

Stanners, R. F., Meunier, G. F., and Headley, D. B. 1969. Reaction time as an index of rehearsal in short term memory. *Journal of Experimental Psychology* 82:566–570.

Sturt, M. 1925. *The psychology of time*. New York: Harcourt.

Swift, E. Y., and McGeoch, J. A. 1925. An experimental study of the perception of filled and empty time. *Journal of Experimental Psychology* 8:240–249.

Talland, G. A., and Quarton, G. C. 1966. The effects of drugs and familiarity on performance in continuous visual search. *Journal of Nervous and Mental Disease* 143:266–274.

Thorndike, E. L., and Lorge, I. 1944. *The teacher's wordbook of 30,000 words.* New York: Teachers College, Columbia University.

von Sturmer, G. 1966. Stimulus variation and sequential judgments of duration. *Quarterly Journal of Experimental Psychology* 18:354–357.

von Sturmer, G. 1968. Time perception, vigilance and decision theory. *Perception and Psychophysics* 3:197–200.

Weiner, H., and Ross, S. 1962a. The effects of "unwanted" signals and d-amphetamine sulfate on observer responses. *Journal of Applied Psychology* 46:135–141.

Weiner, H., and Ross, S. 1962b. Effects of d-amphetamine sulfate on time and brightness perception in human subjects. *Psychopharmacologia* 3:44–50.

Wundt, W. 1874. *Grundzüge der physiologischen Psychologie*. Leipzig: Engelmann.

Yerkes, R. M., and Urban, F. M. 1906. Time-estimation in its relation to sex, age, and physiological rhythms. *Harvard Psychological Studies* 2:405–420.

AMERICAN JOURNAL OF PSYCHOLOGY
September 1977, Vol. 90, No. 3, pp. 447–459

The effects of visual and verbal satiation on a lexical decision task

James H. Neely
Yale University

Immediately before a visually presented target, a string of letters that was to be quickly classified as a word or nonword, the subject saw a prime, a word either semantically related (R) or unrelated (U) to the target. Before this prime, the subject had received visual satiation (V) or both visual and verbal satiation (B) on a word either identical (I) to the prime itself or related to neither (N) the prime or target. Decision times to word targets were faster under condition R than U, but equally so under conditions VI, VN, BI, and BN. Averaged across conditions R and U, decision times to word targets were slower under condition BI than under conditions VI, VN, and BN; decision times to nonword targets under the same four conditions were equal. The results are discussed in terms of the semantic-satiation hypothesis, which they fail to support.

In their thorough review of the literature, Esposito and Pelton (1971) argue that there has been no convincing demonstration of semantic *satiation,* the loss of or decrement in a word's meaning after massed overt verbal repetition or prolonged visual inspection of that word. Of particular relevance to the technique used in the present experiment is a study by Fillenbaum (1964), a study which purportedly demonstrated semantic satiation. Fillenbaum found that when subjects verbally satiated on a word and then decided whether a pair of words were synonymous, decision times were slower if the satiated word was a synonym of one of the words of the test pair than if it was one of the words in the test pair itself. Fillenbaum concluded that satiation on a word in the test pair itself produced a 'sensitization' effect that overrode the semantic satiation effect obtained for the synonyms.

Esposito and Pelton criticize Fillenbaum's conclusion on two grounds. First of all, Esposito and Pelton (1969) closely replicated Fillenbaum's (1964) results in an experiment in which no satiation treatment was given

before the subjects' decisions, and they therefore conclude that satiation played no role in producing Fillenbaum's results. Their second criticism of Fillenbaum's conclusion is that it fails to explain another result obtained by Fillenbaum, namely, why satiation on unrelated words produced longer latencies than satiation on synonyms. They argue that if satiation actually occurs under the conditions of Fillenbaum's experiment, satiation on synonyms should produce slower decision times than satiation on unrelated words.

Unfortunately, the design of Esposito and Pelton's experiment and their interpretation of its results are themselves both open to criticism. Although Esposito and Pelton included nonsatiation conditions in their experiment, they did not include satiation conditions; thus, results under satiation conditions and nonsatiation conditions could not be directly compared. This comparison is important, because there now exists a large body of data showing that when a target follows the presentation of a prime that is identical to it or semantically related to it, the target is processed faster than when it follows a prime that is semantically unrelated to it (see Posner and Snyder, 1975). Such an effect, hereafter referred to as semantic *facilitation,* was also obtained by Fillenbaum (1964) and Esposito and Pelton (1969), since they too found that when the prime given before the test pair was identical or semantically related to one of the words in the test pair, decision times were faster than when the prime was unrelated to the two words in the test pair. The crucial point is that these facilitation effects must be taken into account when one assesses the effects of satiation. That is, contrary to the argument of Esposito and Pelton (1969, 1971), the demonstration that a word has undergone a loss in meaning through satiation does not require that the semantic facilitation effect obtained under nonsatiation conditions be *reversed* under conditions in which the prime has been satiated. Rather, it requires only that the size of the facilitation effect obtained under nonsatiation conditions be *reduced* under conditions in which the prime has been satiated.[1]

Given the present argument that an adequate demonstration of semantic satiation is provided when satiation on the prime reduces the size of the semantic facilitation effect obtained when the prime has not undergone satiation, it is worth noting that when test pairs followed verbal satiation on the prime in Fillenbaum's (1964) study, the facilitation effect of 105 msec (averaged across two experiments) was smaller in magnitude than the facilitation effect of 172 msec that Esposito and Pelton (1969) obtained when no satiation preceded the prime. Since the two studies used the same stimulus materials, it appears that the verbal satiation treatment in Fillenbaum's study reduced the size of the facilitation effect obtained

under nonsatiation conditions in Esposito and Pelton's study by nearly 70 msec, a result in accord with the conclusion that verbal satiation produces a loss of meaning. Of course, the strength of such a conclusion is weakened somewhat because it depends on a comparison of the magnitudes of the semantic facilitation effects obtained in different experiments. Consequently, the results of Fillenbaum (1964) and Esposito and Pelton (1969) cannot be unambiguously interpreted as either demonstrating or disconfirming semantic satiation.

The purpose of the present experiment was to remedy this situation by combining within a single experiment the conditions necessary for drawing a valid inference about whether or not a word loses its meaning through either visual satiation or visual *and* verbal satiation. Rather than using the synonymic decision task employed by Fillenbaum (1964) and Esposito and Pelton (1969), the present experiment utilized a lexical decision task in which the subjects decide as quickly as possible whether or not a string of letters is an English word. Previous work by Meyer and Schvaneveldt (1971) and Schvaneveldt and Meyer (1975) has demonstrated a semantic facilitation effect with such a task, in that a word is more quickly recognized as a word when it follows a semantically related prime than when it follows a semantically unrelated prime. The design of the present experiment allows one to assess the size of this facilitation effect under conditions in which the prime itself has received either prior visual satiation or both visual and verbal satiation and under conditions in which the prime has been preceded by either visual satiation or both visual and verbal satiation on a semantically unrelated word. If semantic satiation occurs, the semantic facilitation effect should be smaller when the prime itself has been satiated than when the prime has been preceded by satiation on an unrelated word.

METHOD

Design

The dependent measure was the amount of time it took each subject to decide whether or not the target (a string of letters) was an English word. Each target was preceded by satiation on a word and then the presentation of a prime.

For the trials in which the target was an English word, and immediately before being presented the prime, which was either semantically related (R) or semantically unrelated (U) to the target, the subjects satiated on a word that was either identical to the prime (I) or semantically related to neither the prime nor target (N). These two factors, the relationship of *prime and target* (R or U) and of *satiated word and prime* (I or N), were crossed with two kinds of

satiation treatment, visual satiation (V) and both visual and verbal satiation (B), to yield eight conditions. Instances of each of the eight are presented in Table 1.

Since one of the primary purposes of the present investigation was to compare the effects of visual satiation with the effects of visual *and* verbal satiation, any nonspecific effects associated with the continuous operation of the articulatory apparatus during the visual and verbal satiation treatment needed to be controlled for. As can be seen in Table 1, these nonspecific effects of overt verbalization were controlled for by requiring that the subjects repeat, during the visual-satiation treatment, a pretrial cue. This cue was a number the experimenter had spoken aloud before the trial began.

Since the relationship of prime and target (R or U) was a pseudodistinction when the targets were nonwords, there were only four conditions here: VI, VN, BI, and BN. Each subject received each of the eight different conditions using words as targets and each of the four different conditions using nonwords as targets.

Stimulus materials and list construction

When the targets were English words, the primes and the targets for the R conditions were taken from an atlas of normative free-association data (Shapiro and Palermo, 1968) and were chosen such that when the prime served as a stimulus, the target was given as the primary response at least 40% of the time. For the U conditions, the primes also came from the norms (either as a primary response or as a stimulus word) but were semantically unrelated to the targets. The to-be-satiated words for the N conditions were drawn from the norms in a similar fashion and were chosen so as to be unrelated to both the primes and the targets.

When the targets were nonwords, they were made from other words drawn from the norms by changing one letter in the word (e.g., BRUSH to GRUSH). The change was such that each nonword target was pronounceable and conformed to the rules of English orthography.

A base list consisting of five blocks was constructed using these materials. All items appeared only once. The first block, the practice block, was given to all subjects and consisted of 16 trials with word targets and 16 trials with nonword targets, trials in which the to-be-satiated word was always unrelated to the prime, which in turn was always unrelated to the target. The two similarly constructed practice trials that were given prior to each test block were also identical for all subjects. Each of the four test blocks contained 36 trials: the 2 practice trials followed by the random presentation of 2 instances of each of the eight conditions using word targets, 4 instances of each of the four conditions using nonword targets, plus 2 instances of a condition in which the to-be-satiated word, the prime, *and* the target were identical.[2]

Seven other lists were constructed from this base list, for a total of eight lists, as follows. The second list and the base list contained an identical order of type of satiation treatment (V or B) for the to-be-satiated words, as well as identical to-be-satiated words and primes. The only difference between these two lists was that the word and nonword targets were exchanged such that a trial with a word target in the base list became a trial with a nonword target in the second list and a trial with a nonword target in the base list became a trial with a word target in the second list. (This involved the introduction

Table 1. Instances of the eight conditions using words as targets, showing any pretrial cue, the (to-be-) satiated word, prime, target, and responses required of the subject

Condition	Pretrial cue[a]	Satiated word (slide)	Subject to say aloud	Prime (slide)	Subject to say aloud	Target (slide)
VIR	'seven'	SAUCER	'saucer, seven, seven, . . .'	SAUCER	'saucer'	CUP
VIU	'thirteen'	DISTANT	'distant, thirteen, thirteen, . . .'	DISTANT	'distant'	BAD
VNR	'two'	CROP	'crop, two, two, . . .'	CHAIR	'chair'	TABLE
VNU	'twelve'	DRAW	'draw, twelve, twelve, . . .'	CONFLICT	'conflict'	EMPTY
BIR		PEPPER	'pepper, pepper, pepper, . . .'	PEPPER	'pepper'	SALT
BIU		CAMP	'camp, camp, camp, . . .'	CAMP	'camp'	DOOR
BNR		PACE	'pace, pace, pace, . . .'	UNCLE	'uncle'	AUNT
BNU		JAIL	'jail, jail, jail, . . .'	MAGIC	'magic'	SMOOTH

[a] Spoken aloud by the experimenter in the appropriate conditions (i.e., those involving only the visual satiation treatment).

of *new* word targets and the reshuffling of the nonword targets that occurred in the base list.) Two more lists were derived from these two lists, respectively, by changing the primes such that when the word targets had been preceded by an unrelated prime, they were now preceded by a related prime, and vice versa. (New to-be-satiated words were introduced when the prime itself had been satiated in the first two lists so that the relationship between the to-be-satiated word and the prime remained the same. When a word unrelated to the prime had been satiated in the first two lists, it remained as the to-be-satiated word.) Four more lists were derived from these four lists, respectively, by changing the to-be-satiated word such that all primes that had themselves been satiated in the first four lists were now preceded by satiation on an unrelated word, and vice versa. (This was done by using the previously nonsatiated primes as to-be-satiated words and reshuffling the to-be-satiated words in the N conditions in the first four lists.)

Finally, these eight lists were replicated such that all to-be-satiated words that had been given the V treatment in the first eight lists were now given the B treatment and vice versa. The net effect of using these 16 different 'lists' was that each word or nonword target occurred in each of its appropriate conditions. Thus, any differences in decision times that might obtain for the different conditions cannot be attributed to item effects.

Procedure

Sixteen Yale graduate students, research assistants, and staff members were paid $4.00 each for their participation in the experimental session, which lasted 1¾ hr. All subjects were native English speakers. Each individually tested subject was read general instructions describing the task and was told that any word presented for his decision would be a common English word, so that the experiment would not be a vocabulary or spelling test. Each subject was told to make fewer than 10% errors and to fixate visually on each slide as long as it remained displayed. Each subject was also told that any two consecutive slides within a trial could be the same.

Each trial consisted of three slides successively rear projected on translucent Mylar near the center of a 12-by-16-cm aperture. The first slide contained the to-be-satiated word and remained on for 11 sec. Prior to each trial the experimenter said either nothing or a number less than a hundred. If nothing was said, the subject repeated the word on the first slide over and over two times per second in time with a metronome until the second slide appeared (B treatment); if a number was spoken, the subject read aloud the word on the first slide once and then started repeating the number over and over two times per second until the second slide appeared (V treatment). The numbers were chosen so that they contained the same number of syllables as the word on the first slide. The second slide remained on for 685 msec and the subject read aloud the word on the second slide once. The third slide contained the target, a string of letters which was presented as soon as the shutter for the second slide closed. The target was displayed until the subject responded by pressing the appropriate key. During the 15-sec intertrial interval, the subject was told if he had made a mistake on the previous trial and the experimenter said a number aloud if the next trial was to involve a V treatment. Slide-display durations were controlled by Lafayette shutters and BRS digibit solid-state

timers, and decision times were measured to the nearest millisecond by a Hunter model 1520 timer.

Each of the 16 subjects received a different one of the 16 'lists' described above. For half of the subjects, the 'word' key was to be pressed by the subject's dominant hand and the 'nonword' key was to be pressed by the subject's nondominant hand; for the other half, the opposite was the case.

RESULTS

Word targets

The mean decision times for correct responses to the word targets and the mean percentages of errors to those targets are displayed in Table 2. The number of observations on which the mean decision times were based ranged from 115 to 127.[3] The most salient aspect of the data for the word targets is that decision times were faster when the word target followed a semantically related prime (R) than when it followed a semantically unrelated prime (U). This semantic facilitation effect was obtained for both satiation treatments under condition N: for treatments V and B there were respective facilitation effects of 25 and 33 msec. However, contrary to what would be predicted if semantic satiation had occurred, the facilitation effects were equally large under condition I: for treatments V and B, respective effects of 22 and 38 msec.

Table 2. Mean decision times (in msec) and errors (%) to the word targets

Satiated word and prime	Type of satiation treatment: Visual (V)	Both visual and verbal (B)
Prime and target related (R)		
Identical (I)		
Time	594	618
Error	1.6	.8
Unrelated (N)		
Time	591	587
Error	.8	.8
Prime and target unrelated (U)		
Identical (I)		
Time	616	657
Error	.8	2.5
Unrelated (N)		
Time	616	620
Error	5.5	1.6

The other salient aspect of the data for the word targets is that the decision times were longer under condition BI (mean, 638 msec, averaged across R and U trials) than under conditions BN, VI, and VN. The latter three conditions produced nearly equal decision times (overall mean, 604 msec).

These general conclusions were supported by a 2 × 2 × 2 × 16 analysis of variance. There was a highly significant effect of the prime's relationship to the target, R or U [$F(1, 15) = 22.33$, $p < .001$, $MS_e = 1238$], and all interactions in which this variable occurred were statistically nonsignificant [all Fs < 1]. Thus, the highly significant semantic facilitation effect was of an equal magnitude for all conditions.

Type of satiation treatment, V or B, did not have a significant main effect on decision times [$F(1, 15) = 2.97$, $p > .10$, $MS_e = 2842$], but the relationship of satiated word and prime did exert a significant influence [$F(1, 15) = 11.73$, $p < .005$, $MS_e = 868$], with the decision times under condition I being longer than those under condition N. However, as noted above, the decision times under condition I were longer than those under condition N only for the B treatment, as revealed by a significant interaction of the relationship of satiated word and prime by type of satiation treatment [$F(1, 15) = 4.70$, $p < .05$, $MS_e = 1{,}801$].

Nonword targets

Table 3 presents the mean decision times for correct responses to the nonword targets and the mean percentages of errors to those targets. The number of observations on which these mean decision times were based ranged from 231 to 243. The means for the nonword targets for the four conditions varied by less than 12 msec.

Thus, unlike the results for the word targets, the decision times for the nonword targets under condition BI were not longer than for those under the other three conditions. The statistical equivalence of the decision times for nonword targets under the various conditions was substantiated by a 2 × 2 × 16 analysis of variance [all Fs < 1].

DISCUSSION

The results of the present experiment failed to demonstrate semantic satiation. That is, compared to prior satiation on words semantically unrelated to the primes (N), prior satiation on the primes themselves (I) did not reduce the magnitude of the semantic facilitation effect obtained when the primes and targets were related (R) rather than unrelated (U).

Table 3. Mean decision times (in msec) and errors (%) to the nonword targets

Satiated word and prime	Type of satiation treatment: Visual (V)	Both visual and verbal (B)
Identical (I)		
Time	740	736
Error	6.0	4.2
Unrelated (N)		
Time	728	735
Error	3.6	4.2

Of course, a proponent of the semantic-satiation hypothesis can offer several reasonable explanations why the present experiment failed to produce evidence of semantic satiation. One possibility is that the lexical decision task does not require that the subject access the meaning of the target. However, James (1975) has provided evidence that when pronounceable nonwords are used as lures, as was the case in the present experiment, the subject does indeed 'look up' the meaning of the target in order to make a lexical decision. Two other possibilities are that the 11-sec satiation period was not long enough or that the rate of verbal repetitions was too slow to produce semantic satiation. Also, since Lackner (1974) has found that subjects experience phonetic distortions of syllables while they listen to recordings of their own repetitions of these syllables but not while they are repeating the syllables aloud to themselves, it may be that subjects in the present experiment did not experience semantic satiation because they were repeating the words aloud to themselves.

However, any argument that the procedures of the present experiment were inadequate to produce a change in the perception of the satiated word must also be able to accommodate the fact that prior verbal satiation on the primes (BI, averaged across R and U) did indeed significantly slow overall decision times to word targets that followed them. In short, given the present negative results and the corpus of negative results reviewed by Esposito and Pelton (1971), it would seem that an acceptance of the semantic-satiation hypothesis should be held in abeyance until well-designed studies provide convincing evidence to support it.

Although the present results were negative with respect to a semantic satiation effect, they clarify why Fillenbaum (1964) obtained a smaller semantic facilitation effect than did Esposito and Pelton (1969). Recall that in Fillenbaum's study the prime itself received prior verbal satiation,

whereas in Esposito and Pelton's study there was no verbalization of any kind given *prior to the prime.* Since the present results show that verbal satiation on the prime itself did not produce a smaller semantic facilitation effect than did verbal satiation on a word semantically unrelated to the prime, one is led to believe that it was the *nonspecific* effects of the overt verbalizations given prior to the primes that reduced the size of the facilitation effect in Fillenbaum's (1964) experiment. Further support for this contention is provided by an experiment (Neely, 1976) using stimulus materials virtually identical to those of the present experiment. Under conditions in which the onset asynchrony between the prime and the target was comparable to the one used in the present experiment, but in which there were no verbalizations of any kind given by the subject prior to the prime, there was a semantic facilitation effect of 51 msec, which is nearly twice as large as the facilitation effect of 29 msec obtained here under conditions in which overt verbalizations were given before the prime.

Of course, it is unfortunate that the present experiment did not include a condition in which no overt verbalizations were required during the satiation period. Such a condition was omitted in the present study because its primary purpose was to compare directly the effects of visual satiation with the effects of visual-plus-verbal satiation and was not to compare directly the effects of verbalization with the effects of no verbalization. Thus, in future research there should be a direct comparison of the sizes of the semantic facilitation effects that are obtained under conditions in which semantically unrelated verbalizations or no verbalizations are given before the primes. If such a direct comparison shows that overt verbalizations reduce the size of the facilitation effect, additional research could examine whether overt verbalizations inhibit the spread of activation between semantically related logogens or inhibit the ability of the subject to direct limited-capacity attention to logogens semantically related to the priming word (Neely, 1977).

Finally, note that although there was no evidence that verbal satiation on the primes reduced the size of the semantic facilitation effect, there was evidence that it slowed overall decision times to word targets but did not affect decision times to nonword targets. Given the currently available data, speculation on the mechanisms underlying the differential effects of prior verbal satiation on the primes upon decision times to word and nonword targets could only be frivolous. Nevertheless, the point does need to be made that verbal satiation on the primes did differentially affect decision times to the word and nonword targets (if one is willing to accept the null hypothesis for the decision times to nonword targets).

Thus, it appears that two separate factors may underlie the processing of word and nonword targets in the lexical decision task.

In summary, the present experiment yielded three main results. First, contrary to the semantic-satiation hypothesis, there was no evidence that verbal satiation on a prime itself selectively interferes with lexical decisions to words semantically related to that prime. Second, when considered in conjunction with data from other experiments, the present data suggest that any kind of overt verbalizations made by the subject before the presentation of the prime act to reduce the size of the semantic facilitation effect. Third, verbal satiation on the prime itself interferes with overall decision times to word targets while leaving decision times to nonword targets unaffected. These three results will need to be incorporated into any theory of the effects of overt verbal repetitions and the mechanisms underlying semantic facilitation in the lexical decision task.

Notes

This research was supported in part by Grant 40310X to Dr. Michael I. Posner from the National Science Foundation and was conducted while Dr. Posner was a Visiting Professor at Yale and the author was a National Science Foundation predoctoral fellow there. Requests for offprints should be sent to the author, who is now at the Department of Psychology, University of South Carolina, Columbia, South Carolina 29208. Received for publication May 27, 1976.

1. In fairness to Esposito and Pelton (1969, 1971), their argument is based on a distinction between two mechanisms by which a word might lose its meaning through verbal satiation. If there is a verbal transformation (see Warren, 1968), as when a subject experiences phonetic distortions while listening to a recording of identical massed repetitions of a word, the word 'Atlantic,' for example, would become meaningless only after the subject began hearing something like 'lanticut.' Thus, the meaning itself actually remains intact but is no longer being contacted by the nonsense syllables 'lanticut.' If, on the other hand, it is the meaning itself that undergoes the change, the subject would begin to hear the word as meaningless gibberish only after it had lost its meaning. Esposito and Pelton further argue that if prior satiation on the prime not only diminishes the magnitude of the semantic facilitation effect but in fact reverses it so that there is a semantic *inhibition* effect (i.e., so that the subject is slower to respond if the target follows a related prime than if it follows an unrelated prime), a strong case could be made for a "true loss of meaning." That is, if the prime merely fails to contact its meaning after satiation, one could argue that the target is following a functional nonword and that the related and unrelated primes become functionally identical such that the magnitude of the facilitation effect would necessarily be zero; whereas if an inhibition effect were obtained, one could argue that while the subject is satiating on a related word such as 'wife,' the lexical item 'husband' is also losing its meaning via semantic generalization, thus making it difficult to access the meaning of

'husband' when it is presented as a target. While it is true that a true loss of meaning would be unambiguously supported if an inhibition effect were obtained after satiation on the prime, it is not true that an inhibition effect *must* be obtained in order to support the argument that a prime has lost its meaning through satiation. A mere reduction in the size of the semantic facilitation effect is sufficient evidence for a loss of meaning; however, it does not permit one to determine whether the loss of meaning was, using Esposito and Pelton's terminology, due to a true loss of meaning or to a verbal transformation. Although such a distinction is interesting and potentially useful, the present paper is concerned only with the logically prior question of whether or not meaning *is* lost through satiation and not with discriminating between the mechanisms that might underlie such a loss.

2. Due to experimenter's error, only half of the subjects received the condition in which the satiated word, prime, and target were all identical. The other half of the subjects received the same targets that appeared in this 'identical' condition, but in the IR conditions. Because of this error and the fact that these items were not counterbalanced across conditions (i.e., they appeared only in the so-called identical or IR conditions for all subjects), the data from these items will not be reported.

3. Decision times were excluded for all those trials on which a subject made an error, for those trials on which the experimenter forgot to say a number for a V trial, and for those trials on which decision times were shorter than 100 msec or longer than 1,800 msec. Data from seven subjects for a total of 15 trials were discarded because of the latter criterion (aberrant decision times). The maximum number of observations if everything had been perfect was 128 for word targets and 256 for nonword targets. The mean percentages of error also exclude trials on which the experimenter made errors and on which the decision times were aberrant as defined above.

References

Esposito, N. J., and Pelton, L. H. 1969. A test of two measures of semantic satiation. *Journal of Verbal Learning and Verbal Behavior* 8:637–644.

Esposito, N. J., and Pelton, L. H. 1971. Review of the measurement of semantic satiation. *Psychological Bulletin* 75:330–346.

Fillenbaum, S. 1964. Semantic satiation and decision latency. *Journal of Experimental Psychology* 68:240–244.

James, C. T. 1975. The role of semantic information in lexical decisions. *Journal of Experimental Psychology: Human Perception and Performance* 1:130–136.

Lackner, J. R. 1974. Speech production: Evidence for corollary-discharge stabilization of perceptual mechanisms. *Perceptual and Motor Skills* 39: 899–902.

Meyer, D. E., and Schvaneveldt, R. W. 1971. Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology* 90:227–234.

Neely, J. H. 1976. Semantic priming and retrieval from lexical memory: Evidence for facilitatory and inhibitory processes. *Memory and Cognition* 4: 648–654.

Neely, J. H. 1977. Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General* in press.

Posner, M. I., and Snyder, C. R. R. 1975. Attention and cognitive control. In *Information processing and cognition: The Loyola symposium,* ed. R. L. Solso. Hillsdale, N.J.: Erlbaum Associates.

Schvaneveldt, R. W., and Meyer, D. E. 1975. Retrieval and comparison processes in semantic memory. In *Attention and performance IV,* ed. S. Kornblum. New York: Academic Press.

Shapiro, S. I., and Palermo, D. S. 1968. An atlas of normative free association data. *Psychonomic Monograph Supplements* 2(12):219–250.

Warren, R. M. 1968. Verbal transformation effect and auditory perceptual mechanisms. *Psychological Bulletin* 70:261–270.

of using the same inducing block for each subject, two inducing blocks, one wider and one narrower than the test block, were alternated day by day. The aims of the experiment were to investigate trends in the measures across days with alternating inducing blocks, the influence of other task features, and finally, the reliability of measures of induction and their relationship with personality variables.

First, the disappearing aftereffect with repeated measurements using the same inducing block may simply be a result of the pre- and postinduction judgments converging on the same asymptote. If so, then stabilizing the pre- and postinduction judgments so that they do not approach the same floor or ceiling should maintain the aftereffect. By alternating the inducing blocks day by day, we hoped to avoid floor and ceiling effects. A stable aftereffect would imply that the diminishing aftereffect of earlier studies was an artifact of floor and ceiling effects on the pre- and postinduction judgments.

Second, certain effects attributed to the standard measurement procedure have been documented. The standard wedge-adjustment procedure involves having subjects make judgments of width to find the point of subjective equality along a tapered wedge. The procedure has these characteristics: there is a tendency to move up the wedge toward the wider end within each set of pre- and postinduction judgments (Weintraub et al., 1973) ; there is a psychophysical error of anticipation, in that subjects stop too soon relative to the starting point on the wedge (see Blitz, Dinnerstein, and Lowenthal, 1966; Costello, 1962; Weintraub et al., 1973) ; and there is, in almost all cases, a positive constant error, whereby mean pre- and postinduction judgments are greater than the actual width of the test block (see Weintraub et al., 1973; Weintraub and Herzog, 1973).

Third, the kinesthetic aftereffect has been used as an indicant of personality (Petrie, 1967; Sales, 1971), although its suitability or usefulness as such has been questioned (Becker, 1960; Brown, 1965; Morgan and Hilgard, 1972; Morgan, Lezard, Prytulak, and Hilgard, 1970; Weintraub et al., 1973). Is the aftereffect a reliable measure of individual differences, and does it correlate with other measures in a theoretically meaningful way? The results have been mixed. Weintraub et al. (1973) examined two measures of individual differences in induction: the *residual change score,* which eliminates the influence of preinduction judgments on the measure of induction, and the less statistically justifiable *aftereffect score,* which is the difference between preinduction and postinduction scores. Only modest test-retest reliability was found, and that only for scores generated by using the wide inducing block. The personality dimension tapped by induction is termed augmentation/reduction, the tendency to

augment incoming stimulation versus the tendency to reduce it (Petrie, 1967). Using a *composite* residual change score with a respectable estimated reliability, Weintraub, Green, and Herzog (1973) found no meaningful pattern of relationships with questionnaire items intended to assess augmenting/reducing tendencies.

Recently, Baker and Mishara (1974) have argued that using a composite measure of induction and assessing its relationship to individual questionnaire items is inappropriate. Instead, one should use only the *first* aftereffect score obtained from each subject and assess its relationship to a composite questionnaire score. They argued that because of differential carry-over effects, measures of induction lack test-retest reliability. However, the first aftereffect score can be shown to be related to other phenomena (e.g., personality variables). Thus, the aftereffect possesses concurrent validity and, therefore, "true test reliability" (Campbell and Fiske, 1959). To demonstrate such relationships, it is important that the other phenomena be reliably measured. Using a measure compounded from several questionnaire items is one way of achieving such reliability. Following this line of reasoning, Baker and Mishara (1974) reanalyzed the data of Weintraub et al. (1973). They used the first-session aftereffect score from the wide inducing block (Weintraub et al. discarded first-session scores in determining a composite residual change score) and formed a single 'reducer' score for each subject by summing across items on the questionnaire. These aftereffect and reducer scores were significantly related for the whole sample, and the relationship was stronger for males than for females. Therefore, a final purpose of the present study was to reexamine the various approaches to the kinesthetic aftereffect as a measure of personality, using a much larger sample than is usually employed.

METHOD

Subjects

The subjects were 71 undergraduates from Grand Valley State Colleges. They served to fulfill part of a requirement of the introductory psychology course.

Apparatus

The apparatus consisted of three wooden blocks, with parallel sides, and a wooden comparison wedge. The wedge was 76.2 cm (30 in.) long and varied in width from 1.27 cm (.5 in.) at the narrow end to 10.16 cm (4.0 in.) at the wide end. The test block was 5.08 cm (2.0 in.) wide and 15.24 cm (6.0 in.) long. Two inducing blocks, 3.81 cm (1.5 in.) and 6.35 cm (2.5 in.) wide, were used on alternate days; each was 15.24 cm long. Blocks and wedge were

equipped with H-shaped riders within which the subject placed thumb and forefinger while making judgments. The test and inducing blocks were always to the left of the standing subject, the comparison wedge always to the right, its wide end forward. The point of objective equality on the wedge (i.e., the point at which its width was equal to the width of the test block) was aligned with the midpoint of the test or inducing block. The apparatus was concealed except during testing, at which time the subject wore a large cardboard collar that prevented his or her seeing the equipment.

Procedure

Each subject served in five separate daily sessions. The first session was always on Thursday, with the four other sessions at the same time of day on the following Monday through Thursday. All subjects judged the 5.08-cm test block. In the first session only, each subject felt the full length of the wedge and then made two practice judgments of the test block. This was followed by four baseline, or preinduction, judgments of the test block; a 60-sec induction period with the wide inducing block; four postinduction judgments of the test block; a 60-sec induction period with the narrow inducing block; and four postinduction judgments of the test block. For the four remaining sessions, the procedure consisted of four preinduction judgments; a 60-sec induction period; and four postinduction judgments. The wide inducing block was used in the second and fourth sessions, the narrow inducing block in the third and fifth sessions.

For each judgment, the subject held the test block with the thumb and forefinger of the left hand and simultaneously moved the thumb and forefinger of the right hand along the wedge to find the point at which the width of the wedge felt equal to the width of the test block. The subject was allowed to move the right hand back and forth on the wedge to 'zero in' on this point. Between judgments, both hands were at the subject's sides. For both the pre- and postinduction judgments, half of the judgments started from a point on the wedge that was 32.0 cm (12.6 in.) away from the point of objective equality and toward the narrow end of the wedge (ascending judgments, A); the other half started from a point the same distance away from the point of objective equality but toward the wide end of the wedge (descending judgments, D). Order of ascending and descending judgments was balanced (ADDA) for each subject. During the 60-sec induction period, each subject rubbed the appropriate inducing block with the thumb and forefinger of the left hand at the rate of approximately one back-and-forth movement per second. The right hand was at the subject's side during this period. The aftereffect score for each subject was determined for each session by subtracting the mean of the four preinduction judgments from the mean of the four postinduction judgments.

At the end of the last session, each subject was given a 25-item questionnaire to answer. Data on the questionnaire are available for 67 of the subjects. The items dealt with phenomena that had been found to relate to the aftereffect score or could reasonably be expected to be associated with an augmentation/reduction dimension, topics such as birth order, childhood environment, personal habits and activities (e.g., drinking, smoking, sleeping, partying, fingernail biting, studying), pain tolerance, and friendships.

RESULTS

Trends across days

The top of Figure 1 presents the mean pre- and postinduction judgments as a function of testing day and type of judgment. The dashed line in the figure traces the temporal sequence of the judgments, connecting them in their testing order. Note the trends. Rubbing the wide inducing block led to a decrease in judged width as compared with the just-previous set of judgments. Conversely, rubbing the narrow inducing block led to an increase. Finally, merely judging width on the wedge without any in-

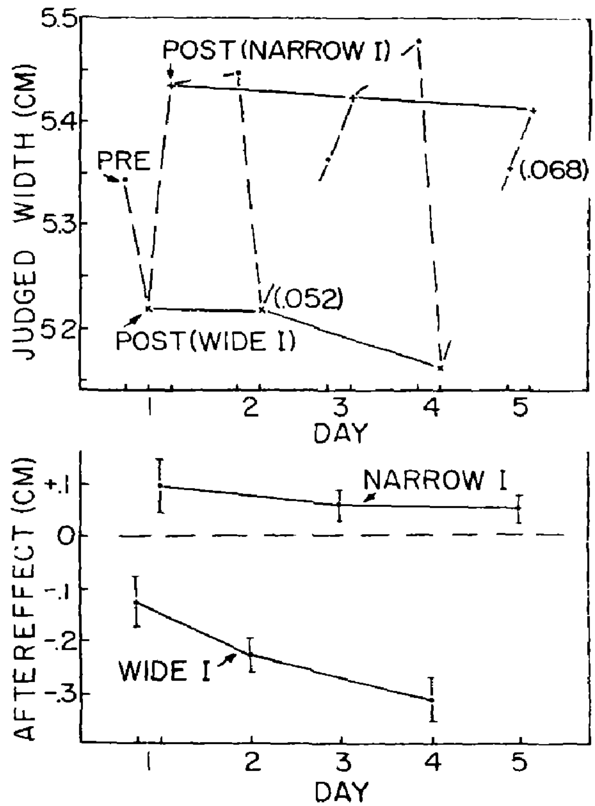


Figure 1. At the top are the mean widths for preinduction (Pre) judgments and for postinduction judgments preceded by the narrow (Post/Narrow) and wide (Post/Wide) inducing blocks; the dashed line connects the means in proper temporal order and the smallest and largest $s_{\bar{X}}$ are given in parentheses. At the bottom are the mean aftereffect scores, with ± 1 $s_{\bar{X}}$ depicted around each post- minus preinduction difference for the narrow and wide inducing blocks

tervening rubbing led to increases in the judged width of the test block, as the elevation of preinduction judgments over postinduction judgments on the preceding day indicates.

The traditional measure of the kinesthetic aftereffect subtracts preinduction scores taken on any day from postinduction scores for that day. These aftereffect (difference) scores are plotted at the bottom of Figure 1. The aftereffect means showed the traditional contrast effect, negative for the wide inducing block and positive for the narrow inducing block.

Day 1 was initially planned as a practice day, unique because (a) it was separated by four days from the rest of the series, in the hope that long-term effects would dissipate, and (b) it involved a preinduction measure followed by two induction experiences, one with each inducing block. The aftereffect associated with the narrow inducing block on day 1 might have been measured differently, treating the postinduction judgments from the wide inducing block as the measure of the true prior state of the organism. If so, then the aftereffect on day 1 for the narrow inducing block would have been more than double its plotted value. However, aftereffects are traditionally treated as a difference between preinduction and postinduction scores, and statistical analysis of the data conformed to this procedure.

Repeated-measures analyses of variance were carried out separately on the data for the wide and narrow inducing blocks. For the *wide* inducing block, the main effect of judgment, pre- or postinduction, was significant [$F(1, 70) = 54.71$, $p < .001$]; the interaction of judgment by day, day 1, 2, or 4, was also significant [$F(2, 140) = 6.83$, $p < .005$]. Post hoc analysis of the interaction using the Tukey *B* test (Wike, 1971) with Cicchetti's (1972) modification for interaction tables was carried out. It revealed that the postinduction means differed significantly from the preinduction means on all three days (days 1, 2, and 4), that the day 1 preinduction mean differed from the other two preinduction means, and that the postinduction means did not differ [$s_{\bar{x}} = .023$; $p < .05$ for all significant effects]. In other words, the interaction was caused by the preinduction mean on day 1. Beyond day 1, the pre- and postinduction means (and, therefore, the mean aftereffects) were stable. For the *narrow* inducing block, only the main effect of judgment was significant [$F(1, 70) = 7.75$, $p < .01$]. Therefore, the pre- and postinduction means were stable and differed significantly from each other by a constant amount on all three testing days (days 1, 3, and 5). Clearly, with inducing blocks alternated day by day, there were significant aftereffects for both inducing blocks and no tendency for the aftereffects to decline toward zero with repeated testing.

Task features

First, there was evidence of a rise within each set of four pre- and four postinduction judgments. For each day, the mean of the first judgments within each set was subtracted from the mean of the fourth judgments (the first and fourth judgments of a foursome were ascending judgments). In 10 of 11 cases, the difference was positive (statistically significant by the binomial test), evidence for a rise in judgments within each set. Likewise, the mean of the third judgments within each set was greater than the mean of the second judgments (both descending judgments) in 10 of 11 cases. For convenience, we refer to this upward tendency in wedge judgments as the *wedge-judgment bias.*

Second, the typical error of anticipation occurred within each set of judgments. This was evaluated by subtracting the mean of the ascending judgments from the mean of the descending judgments for each of the 11 sets of judgments. In all cases, the difference was positive, indicating once again that subjects tended to stop before reaching the point of subjective equality found by averaging ascending and descending judgments.

Third, the constant error was positive, as usual. As can be clearly seen in Figure 1, all pre- and postinduction means were widths much greater than the point of objective equality (5.08 cm).

Fourth, as the lower panel of Figure 1 shows, the absolute size of the mean aftereffect was always greater with the wide inducing block than with the narrow one. A previous demonstration of this outcome by Weintraub, Green, and Herzog (1973) involved a between-groups rather than a within-groups comparison.

Reliability

Table 1 presents the correlation coefficients among measures both within and between days. It summarizes the correlations for all combinations of inducing blocks. Preinduction scores were highly intercorrelated, postinduction scores were highly intercorrelated, and the two types of scores correlated highly with one another, especially for measures taken on the same day.

In studying individual differences in induction, it is desirable to deal realistically with the initial bias of subjects as assessed by their preinduction scores. The correlations between preinduction and postinduction scores on the same day (Table 1) reveal the carry-over that must be accounted for in order not to swamp a measure of induction with extraneous variance. Residual change scores were calculated as an improved procedure for assessing individual differences in induction. A residual change

Table 1. Median product-moment correlations (r) for preinduction, postinduction, residual change, and aftereffect scores for all combinations of inducing blocks

	Inducing block(s)		
	Same block		
	Wide block	Narrow block	Different blocks
Same day			
Preinduction and postinduction	.79	.86	
Preinduction and residual change	.00	.00	
Postinduction and residual change	.62	.51	.38(2)
Postinduction and postinduction			67(1)
Residual change and residual change			.48(1)
Preinduction and aftereffect	−.45	−.30	
Postinduction and aftereffect	.18	.33	.20(2)
Aftereffect and aftereffect			.54(1)
Aftereffect and residual change	*.89*	*.95*	*.46(2)*
Different days			
Preinduction and postinduction	.58(6)	.64(6)	.71(12)
Preinduction and residual change	.14(6)	.16(6)	.23(12)
Postinduction and residual change	.33(6)	.30(6)	22(16)
Preinduction and preinduction	.70	.63	.76(4)
Postinduction and postinduction	.65	.67	.66(8)
Residual change and residual change	.31	.17	.12(8)
Preinduction and aftereffect	−.18(6)	−.02(6)	−.12(12)
Postinduction and aftereffect	.00(6)	.11(6)	−.09(16)
Aftereffect and aftereffect	.25	.08	.12(8)
Aftereffect and residual change	.23(6)	.10(6)	.07(16)

Note: In parentheses is given the number of rs on which the median is based whenever the number is not 3. $N = 71$ for each r; $r = \pm .24$, $p < .05$, two-tailed.

score is the deviation of a postinduction score from its predicted value based on linear regression of postinduction on preinduction scores. (See a slightly more extended treatment and justification in Weintraub et al., 1973.) Residual change scores are, by design, uncorrelated with the preinduction scores on which they are based. Note in Table 1 the zero correlations between preinduction and residual change scores on the same day. Thus, residual change scores do account for the subjects' initial biases as measured by their preinduction scores. The traditional aftereffect difference measures, called aftereffect scores, appear in Table 1 for purposes of comparison. All such difference scores share the undesirable property of

correlating negatively with the measure that is supposed to be eliminated from the data by the subtraction process. As shown in Table 1, the correlations between same-day preinduction and aftereffect scores were significantly negative.

Although the results will be discussed in terms of residual change scores, one may follow the arguments via the aftereffect (difference) scores. The two types of measures correlated highly with one another when based on the same pair of scores (see correlations for aftereffect and residual change scores on the same day). In addition, the two measures correlated about as well with one another across days as they did among themselves.

Reliability across days was modest but consistently significant for residual change scores based on data from the wide inducing block (see correlations for residual change and residual change scores on different days for the wide inducing block). For the narrow inducing block, the correlations were generally smaller, but all three were positive. Correlations between sets of residual change scores for *different* inducing blocks across days also tended to be positive but small. Hence, reliability was best for the wide-block scores, but there was some tendency for subjects to maintain their positions in the distribution of scores even when inducing blocks were alternated across days. It should be emphasized that when pre- and postinduction judgments are highly correlated, as they are, there cannot be a great deal of variance left to be captured in the residual change score.

Following an earlier strategy (Weintraub et al., 1973), we decided to build more reliable measures of individual differences by forming composite scores. For each subject, composite scores were derived by averaging scores from relevant experimental sessions. Three composite residual change scores were computed: one from sessions involving the wide inducing block, one from those with the narrow inducing block, and one based on the data from both blocks. Composite preinduction and aftereffect scores were derived in a similar manner. All composite scores were intercorrelated. Two features of the data deserve comment. First, the composite preinduction scores provided a far more reliable measure of individual differences than either of the composite induction scores (residual change or aftereffect), just as the correlations in Table 1 between unaggregated scores would indicate. Moreover, the reliability of preinduction scores was not in the least diminished by alternating inducing blocks across days [$r = .93$ between composite preinduction scores based on the wide versus the narrow inducing blocks]. Second, the composite residual change scores make clear what is only suggested in Table 1: there was a modest but significant tendency for subjects to maintain their positions in the distribution of residual change scores even though inducing blocks were alternated

across days [$r = .39$ between composite residual change scores based on the wide versus the narrow inducing blocks].

Personality correlates

The relationship between questionnaire items and induction was examined. Scores on one of the measures of induction (measures tended to be correlated with one another) were correlated with the responses to each item on the questionnaire. Product-moment correlations were computed for each of the 12 questionnaire items having four or more ordered response alternatives. Biserial correlations were computed for each of the remaining 13 two-alternative items. As a typical example, correlations using the composite residual change score based on both inducing blocks produced 3 items that reached statistical significance [$p < .05$]: reducers, so judged on the basis of their induction scores, reported less fear of getting an injection and a greater preference for watching car races than augmenters, and they were more likely to be first-born or only children than augmenters. This and the similar outcomes involving the other measures of induction cannot be considered encouraging because the entire 25-item questionnaire was designed to assess augmentation/reduction. Thus, 1¼ items would achieve significance by chance alone.

Since Baker and Mishara (1974) had unearthed statistically significant findings in the data of Weintraub et al. (1973), the data reported herein were also analyzed following their procedures exactly. This analysis involved looking at the residual change scores for only the wide inducing block on day 1, treating sex of subject as a variable, and increasing the reliability of the questionnaire data by forming a composite questionnaire score for each subject. Neither Baker and Mishara's treatments nor any amalgam of their approach and ours produced noteworthy outcomes.

Since preinduction scores proved so reliable, a check was made to determine if they were better predictors of questionnaire data than were induction measures. Best predictability of the composite questionnaire score was associated with the wide-block composite preinduction score, but the correlations [$r = .20$ for males, .16 for females, .18 for all subjects] were not statistically significant.

DISCUSSION

Artifacts and induction

Three sources of artifact in the measurement of the kinesthetic aftereffect were documented in this study. One is the classical error of antici-

pation: subjects stop at a narrower width when ascending the wedge (moving toward the wide end) and at a wider width when descending. This source of error can be controlled by using both ascending and descending trials.

A more troublesome artifact is the wedge-judgment bias, the tendency for judgments to move toward the wide end of the wedge within each set of judgments. This artifact cannot be counterbalanced across judgments, and it leads to two testable consequences. First, the point of subjective equality (PSE) computed from the very first set of preinduction judgments should be too wide. It is [PSE = 5.34 cm, $s_{\bar{x}}$ = .053 cm, compared to the actual width of 5.08 cm]. Weintraub and Herzog (1973) found that when the wedge was not used, the initial error in the PSE disappeared. Second, if rubbing does not intervene, a set of judgments will be wider on the average than a just-prior set. When preinduction judgments on one day in this study are compared with the postinduction judgments of the previous day, one sees that the increase in mean width occurred in every case (Figure 1). This finding cannot be explained by the well-documented dissipation of induction with time (e.g., Bakan, Myers, and Schoonard, 1962; Bourne, Kepros, and Beier, 1963), because that would result in a *decrease* in preinduction means on days following the narrow inducing block. An increase in mean width was also obtained by Weintraub et al. (1973), especially among subjects who made wedge judgments but never rubbed an inducing block. Hence, the wedge-judgment bias should be considered in interpreting findings whenever the wedge is used. The bias can be eliminated by eliminating the wedge, as Weintraub and Herzog (1973) did.

A third source of artifact involves the ceiling and floor effects that can occur following repeated induction. With repeated induction using the same inducing block, these effects are especially noticeable (Weintraub et al., 1973). In the present study, with alternating inducing blocks, a ceiling effect seems evident for wide-block preinduction judgments (Figure 1); a floor effect may exist. Finger span may account for the ceiling effect, in that a comfortable span between thumb and forefinger may be far smaller than one might believe. The larger aftereffects with the wide inducing block can be explained on the assumption that the set of blocks have widths closer to the psychological ceiling than to the floor, permitting more room for postinduction change following wide-block induction. An informative experiment would maintain the same differences in width among blocks while reducing all widths by at least 2.5 cm.

It should be emphasized that, despite artifacts, induction is clearly at work. The traditional contrast effect is evident in the mean aftereffects

of Figure 1. Moreover, the experiment demonstrated, as had been hoped, that alternating inducing blocks prevents the measured aftereffect from declining precipitously toward zero, as it had in previous experiments employing repeated induction. Therefore, subjects do not become immune to the effects of rubbing via processes like learning or adaptation. Rather, induction produces predictable effects that are, in turn, constrained in predictable ways by the artifacts discussed above.

Individual differences

We have been able to construct a composite measure of induction that has modest but significant reliability. Neither it nor any of its components exhibits meaningful correlations with the questionnaire as a whole. The questionnaire may be at fault, although its items do have face validity and have yielded significant results for others. Nonetheless, given the imprecision of measurement of both the effects of induction and the variables involved in personality, the prospect of obtaining a consistent link between them seems remote.

Regardless of their importance for personality theory, consistent individual differences in induction are of interest in their own right. The positive correlation between composite wide-block and narrow-block residual change scores means that those individual differences should not be interpreted in terms of a 'susceptibility to rubbing.' Susceptibility would cause a negative correlation, because susceptible subjects would show a large positive aftereffect following narrow-block induction and a large negative aftereffect following wide-block induction. Rather, the dimension is augmentation/reduction. An augmenter readily increases the magnitude of judgments following narrow-block induction but only reluctantly decreases them following wide-block induction. A reducer readily decreases judgments but only grudgingly increases them.

Reliable induction effects should lead augmenters and reducers to move apart in the distribution of wedge-judgment scores with repeated measurement. Three predictions follow. First, the variability of pre- and post-induction scores should increase from the first to the last session. Second, augmenters and reducers should have significantly different postinduction scores on the final day of testing. Both predictions received some support in an earlier study (Weintraub et al., 1973). In the present study, variances of preinduction scores increased monotonically from the first to the last session [first- and last-session variances differed significantly, $t(69) = 3.01$, $p < .005$, one-tailed], but variances of postinduction scores did not change. As for the second prediction, subjects classified as augmenters and

reducers by a median split of the wide-block composite residual change scores had significantly different postinduction scores on the final day of testing [$z = 2.64$, $p < .005$, one-tailed Mann-Whitney U test]. A third prediction is that compared to preinduction scores, the variances of postinduction scores on any day should be larger. Only day 1 conformed to this prediction. These mixed results underscore the modest reliability of measures of induction and suggest caution in making any conclusions about individual differences in induction.

In conclusion, then, alternating the width of the inducing block will serve to preserve the kinesthetic aftereffect. Individual differences in response to induction can best be characterized as lying along an augmentation/reduction dimension. We close with the caveat that one must be alert for lurking artifacts: it is difficult to know if one has the real tiger by the tail. It is at least conceivable that the most profound individual differences are associated with the wedge-judgment bias and that these differences are the primary cause of the substantial correlation to be found between any two sets of preinduction and/or postinduction scores.

Notes

The authors thank Mike Cahill for running subjects, Russ Knapp for assisting in data analysis, and A. Harvey Baker and Irene W. Kostin for making available to them the scoring system for the personality questionnnaire used by Baker and Mishara (1974). Preparation of this manuscript was supported in part by National Institute of Mental Health Grant MH23767 to the second author. Offprint requests should be addressed to Thomas Herzog, Department of Psychology, Grand Valley State Colleges, Allendale, Michigan 49401. Received for publication June 29, 1976; revision, September 27, 1976.

References

Bakan, P., Myers, L. B., and Schoonard, J. 1962. Kinesthetic after-effects and length of inspection period. *American Journal of Psychology* 75:457–461.

Baker, A. H., and Mishara, B. 1974. Kinesthetic figural aftereffects: Can this simple tactual test really predict personality attributes? Paper presented at the meeting of the Eastern Psychological Association.

Becker, W. C. 1960. Cortical inhibition and extraversion-introversion. *Journal of Abnormal and Social Psychology* 61:52–66.

Blitz, B., Dinnerstein, A. J., and Lowenthal, M. 1966. Relationship between pain tolerance and kinesthetic size judgment. *Perceptual and Motor Skills* 22:463–469.

Bourne, L. E., Jr., Kepros, P. G., and Beier, E. G. 1963. Effect of post-inspec-

tion delay upon kinesthetic figural aftereffects. *Journal of General Psychology* 68:37–42.

Brown, R. A. 1965. Work decrement, kinesthetic aftereffect, and personality. *Journal of Personality and Social Psychology* 2:868–872.

Campbell, D. T., and Fiske, D. W. 1959. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin* 56:81–105.

Cicchetti, D. V. 1972. Extension of multiple-range tests to interaction tables in the analysis of variance: A rapid approximate solution. *Psychological Bulletin* 77:405–408.

Costello, C. G. 1961. Constant errors in the measurement of kinesthetic figural after-effects. *American Journal of Psychology* 74:473–474.

Köhler, W., and Dinnerstein, D. 1947. Figural after-effects in kinesthesis. In *Miscellanea psychologia Albert Michotte*. Paris: Librairie Philosophique.

Morgan, A. H., and Hilgard, E. R. 1972. The lack of retest reliability for individual differences in the kinesthetic aftereffect. *Educational and Psychological Measurement* 32:871–878.

Morgan, A. H., Lezard, F., Prytulak, S., and Hilgard, E. R. 1970. Augmenters, reducers, and their reaction to cold pressor pain in waking and suggested hypnotic analgesia. *Journal of Personality and Social Psychology* 16:5–11.

Petrie, A. 1967. *Individuality in pain and suffering*. Chicago: University of Chicago Press.

Sales, S. M. 1971. Need for stimulation as a factor in social behavior. *Journal of Personality and Social Psychology* 19:124–134.

Weintraub, D. J., Green, G. S., and Herzog, T. R. 1973. Kinesthetic aftereffects day by day: Trends, task features, reliable individual differences. *American Journal of Psychology* 86:827–844.

Weintraub, D. J., and Herzog, T. R. 1973. The kinesthetic aftereffect: Ritual versus requisites. *American Journal of Psychology* 86:407–423.

Wike, E. L. 1971. *Data analysis*. Chicago: Aldine-Atherton.

AMERICAN JOURNAL OF PSYCHOLOGY
September 1977, Vol. 90, No. 3, pp. 475–479

Long-term recency effects in free recall

James W. Aldridge and Michael T. Farrell
Thomas More College

Although Tzeng (1973) and Bjork and Whitten (1974) have obtained positive recency effects in free recall using a procedure designed to eliminate any component of short-term storage, their procedures may not have truly cleared short-term storage. In the present experiment, subjects were presented with four lists for free recall, each list composed of seven sets of noun triples. Presentation was either visual or auditory, and subjects counted backward by sevens before and after each item. Even with short-term storage thus cleared, recency effects were obtained equally for both modalities. No effect of serial position was found in a final test of free recall.

It has generally been considered that the pronounced positive recency effect normally found in the free recall of unrelated items results from readout of the last few list items from a short-term store, distinct from a longer-term store in which the other items are retained (e.g., Glanzer, 1972). Perhaps the strongest evidence for this interpretation is the general finding that when the interval between list presentation and recall is filled with another activity, the recency effect disappears while the rest of the serial-position curve remains unchanged (e.g., Postman and Phillips, 1965).

Tzeng (1973) and Bjork and Whitten (1974) have recently found recency effects, however, under paradigms intended to eliminate any component of short-term storage. Instead of presenting an entire list of items followed by a distractor activity, these investigators required subjects to count backward (Tzeng) or to engage in arithmetic activity (Bjork and Whitten) before and after each item. This was intended to clear short-term storage after each item's presentation, leaving only long-term storage to produce any serial-position effects. Clear effects of both primacy and recency were obtained in both studies.

The obvious interpretation of these studies would seem to be that invocation of a short-term store is unnecessary to explain recency effects in

free recall. If short-term storage was effectively cleared in the study b Tzeng and the one by Bjork and Whitten, the recency effects obtainec were clearly a long-term phenomenon. Unfortunately, there are two rea sons for believing that the paradigms used did not effectively clear short term storage. For one thing, the items on the list were single words ir Tzeng's experiment and pairs of words in Bjork and Whitten's. A crucia assumption in these studies is that the distractor activity cleared each ol these items from short-term storage. Murdock (1961), however, found that distraction produces almost no forgetting of a single word over intervals of several seconds. Pairs of words might well be more subject to forgetting, but to assume complete clearing from short-term storage is clearly untenable.

A second reason for believing that the procedures did not clear short-term storage is that both experiments used visual presentation. Several studies (e.g., Kroll, Parks, Parkinson, Bieber, and Johnson, 1970; Parkinson, 1972) have shown that visual presentation leads to greater resistance to subsequent distraction than auditory presentation.

Although the results of Tzeng (1973) and Bjork and Whitten (1974) are potentially quite important, if a long-term recency effect is to be accepted, it must be found in a situation in which short-term storage is truly cleared. The purpose of the present experiment was to determine whether positive recency would be obtained in such a situation. The lists for free recall were constructed using three nouns as each item. Before and after each item, subjects counted backward by sevens for 20 sec. In one condition, visual presentation was used; in a second condition, auditory presentation. It was believed that a recency effect obtained under these conditions, particularly with auditory presentation, could be safely attributed to long-term storage.

METHOD

Subjects

The subjects were 4 introductory psychology students at Thomas More College, who received course credit for their participation; 8 upper-level psychology majors from Thomas More College; and 28 student nurses at Good Samaritan Hospital, Cincinnati, Ohio.

Apparatus and material

The lists for free recall were constructed from triples of common, four-letter monosyllabic nouns. Visual presentation was accomplished using a Kodak Carousel slide projector. Auditory presentation was spoken. Timing was controlled by a tape recording with tones occurring at appropriate intervals. A different set of five lists of seven triples each was used for each subject. The same pool of 35

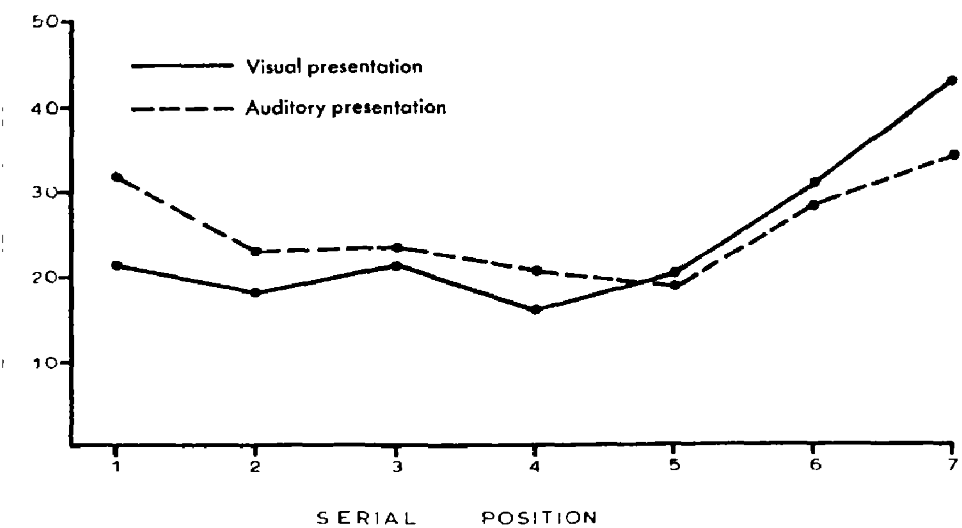
iples was used in constructing each set of lists. Triples were never repeated ithin a set of five lists.

rocedure

Subjects were divided into two groups of 20, one group receiving visual pres- ıtation and one group auditory presentation. Each list began with the tape- corded word "Start." The experimenter then called out a three-digit number. ıbjects were instructed to repeat the number and immediately begin counting ıckward by sevens. After 20 sec, a tone signaled the subject to stop counting ıd the experimenter to present an item. For visual presentation, the triple was splayed for 3 sec until a second tone signaled the experimenter to turn off the ide and call out another number. For auditory presentation, the words were ›oken by the experimenter during the 3-sec interval. Each list consisted of seven iples preceded and followed by 20 sec of counting. After each list, 2 min were lowed for written recall, followed immediately by the next list. After presenta- on and recall of all five lists, there was a 2-min phony debriefing session. Finally, ithout prior warning, subjects were asked to recall all items in all lists in 2 min. During presentation, subjects were instructed to concentrate on the item being resented, to the exclusion of other items in the list.

.ESULTS

The first list for each subject was considered practice and not scored. ince subjects rarely recalled an entire triple, recall was calculated as the ercentage of total words recalled. Mean initial recall scores are shown s a function of serial position in Figure 1. Analysis of variance yielded a gnificant serial-position effect [$F(6, 228) = 8.63, p < .001$], no effect



igure 1. Mean percent recalled in initial recall as a function of serial position

of group [$F < 1$], and no interaction of serial position by group [$F(6, 228) = 1.67, p > .05$]. Mean number of words recalled as a function of serial position in final recall is shown in Figure 2. Items from the first list were again not counted. In final recall, there were no effects of serial position [$F(6, 228) = 1.62, p > .05$] or group [$F(1, 38) = 1.14, p > .05$] and no interaction of serial position by group [$F(6, 228) = 1.81, p > .05$].

DISCUSSION

The results clearly support the interpretation that solely long-term recency effects can be obtained in delayed free recall. Even with three-word items to remember and 20 sec of counting backward by sevens before and after each presentation, rather striking recency effects were found in both the visual and auditory modalities. There was, further, no real indication of an interaction of serial position by modality. It is difficult to see how short-term storage could have contributed to the present results.

The failure to find any serial-position effect in final free recall replicates that of Bjork and Whitten (1974), although Tzeng (1973) found both recency and primacy in final as well as initial recall. Tzeng's subjects, however, had only a total of 40 words to recall at final recall (single words, 10 positions, four lists), while in the present experiment subjects had 105 words and in Bjork and Whitten's study they had 80 words. Tzeng plausibly suggests that having too many items to remember at final recall may lead to a loss of serial-position effects via output interference. If such an interpretation is correct, it suggests that the serial-position effect obtained

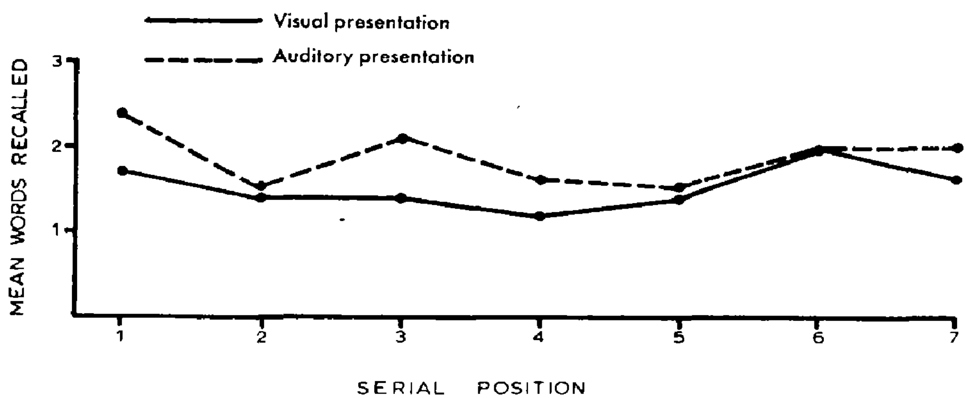


Figure 2. Mean number of words recalled in final recall as a function of serial position

under the present paradigm represents a phenomenon of retrieval, rather than of storage.

As a final point, little should probably be made of the failure to find a primacy effect in initial recall in the present experiments, although such an effect was found by both Tzeng (1973) and Bjork and Whitten (1974). In order to maintain a total number of words per list comparable to that used by Bjork and Whitten, only seven serial positions were used in the present experiment. The lists were therefore probably too short to allow a primacy effect to develop. Bjork and Whitten found, in fact, that the magnitude of their primacy effect increased with the length of the list.

Notes

Requests for offprints should be sent to James W. Aldridge, Department of Behavioral Sciences, Pan American University, Edinburg, Texas 78539. Received for publication October 11, 1976.

References

Bjork, R. A., and Whitten, W. B. 1974. Recency-sensitive retrieval processes in long-term free recall. *Cognitive Psychology* 6:173–189.

Glanzer, M. 1972. Storage mechanisms in recall. In *The psychology of learning and motivation*, vol. 5, ed. G. H. Bower. New York: Academic Press.

Kroll, N. E. A., Parks, T., Parkinson, S. R., Bieber, S. L., and Johnson, A. L. 1970. Short-term memory while shadowing: Recall of visually and of aurally presented letters. *Journal of Experimental Psychology* 85:220–224.

Murdock, B. B., Jr. 1961. The retention of individual items. *Journal of Experimental Psychology* 62:618–625.

Parkinson, S. R. 1972. Short-term memory while shadowing: Multiple-item recall of visually and of aurally presented letters. *Journal of Experimental Psychology* 92:256–265.

Postman, L., and Phillips, L. W. 1965. Short-term temporal change in free recall. *Quarterly Journal of Experimental Psychology* 56:413–419.

Tzeng, O. J. L. 1973. Positive recency effect in a delayed free recall. *Journal of Verbal Learning and Verbal Behavior* 12:436–439.

AMERICAN JOURNAL OF PSYCHOLOGY
September 1977, Vol. 90, No. 3, pp. 481–488

False recognition of synonyms of right and wrong items in multiple-item recognition learning

Donald H. Kausler, David M. Kleim, and Robert M. Yadrick
University of Missouri — Columbia and Columbia College

The hypothesis that right items are processed at a featurally deeper level than wrong items was tested. After one, two, or four study trials on a list of paired right and wrong items, subjects were asked to decide whether individual items, including new ones that were synonyms of the prior ones or unrelated to them, were old or new and, if old, had been right or wrong. A false-recognition effect was found for synonyms of both right and wrong items, suggesting that wrong items, as well as right, are processed at a semantic level. However, that effect took more study trials to appear for synonyms of wrong items, reflecting the disparity between right and wrong items in the rates at which their features accrue frequency units.

Implicit in Ekstrand, Wallace, and Underwood's (1966) frequency theory of verbal-discrimination learning — that is to say, multiple-item recognition learning — is the assumption that the processing of right items does not differ qualitatively from the processing of wrong items. Processing consists of representational and/or rehearsal responses to items whether they function as right or wrong in the pairs. That is, the processing of right items is distinguishable from the processing of wrong items only in terms of the greater number of responses elicited by right items. It is intuitively appealing, however, to view right items as being characteristically processed at a deeper featural level (Craik and Lockhart, 1972) than wrong items. A right item is clearly identified to the subject as the target, the source of information about the pair. As such, it is likely to be processed by means of elaborative rehearsal, an activity directed toward the extraction and frequency tagging of the semantic features indigenous to that item's memorial representation. On the other hand, a wrong item, as a distracting source of information about the pair, may be processed at a shallow level, one that yields little, if any, extraction and frequency tagging of that item's semantic features.

In an earlier study, Kausler, Pavur, and Yadrick (1975) obtained evi-

dence in support of the hypothesis that right and wrong items differ qualitatively in their processing. The procedure called for testing individual items in terms of their oldness (i.e., presence or absence in the study list) and, if identified as old, their function (i.e., right or wrong) on the study list, which consisted of unrelated paired right and wrong items. Kausler et al. included on the test list new items that were homophones of items that had been right and wrong on the study list, as well as control items that were unrelated to the study items. They found more false recognition (Underwood, 1965) of the homophones of right items, but not of the homophones of wrong items, both as compared to the control items. The implication was that the sensory features of right items, but not of wrong items, are extracted and tagged with frequency units during study. However, only one study trial preceded the test in the experiment. Wrong items may receive the same featural processing as right items, but their features may simply accrue frequency units at a slower rate than the features of right items, in agreement with the differential rate of responding proposed by frequency theory. If true, more study trials would be required to produce false recognition of new items that are featurally related to prior wrong items than of new items that are featurally related to prior right items.

The present study replicated the procedure of Kausler et al. (1975) as a means of testing further the hypothesis that the processing of wrong items differs qualitatively, as well as quantitatively, from the processing of right items. However, synonyms, rather than homophones, of prior right and wrong items served as experimental test items. In addition, the number of study trials preceding the single test trial varied from one to four. If wrong items are processed at a shallow level, their semantic features are unlikely to accrue sufficient frequency units to evoke false recognition of their synonyms even after multiple study trials. On the other hand, if the semantic features of wrong items simply accrue frequency units at a relatively slow rate, they should eventually accrue sufficient units to evoke false recognition. By contrast, the rapid accrual of frequency units to the semantic features of right items should quickly promote false recognition of the synonyms of these items, perhaps even after a single study trial.

METHOD

Design and subjects

The overall design was a 3 × 3 mixed factorial. The within-group factor represented three types of relationships, two experimental and one control, between

new test items and prior study items. For one experimental condition (E_R), the test items were synonyms of prior right items; for the other experimental condition (E_W), the test items were synonyms of prior wrong items. For the control condition (C), the test items were words that were unrelated to prior study items.

The between-group factor represented three levels of practice: one, two, or four study trials. The 120 subjects, volunteers from introductory psychology classs who fulfilled a course requirement by their participation, were assigned in randomized blocks to the three levels of practice (each group's $N = 40$). None of the subjects had participated in an experiment on recognition learning before. Each of these three main groups was divided via randomized block assignments into eight subgroups of 5 subjects each. The eight subgroups were needed to insure that each synonym occurred equally often as E_R, E_W, and C items (to be described later).

Lists

The study list for a given subject contained 24 pairs of items, a right item and a wrong one, of which 2 were control pairs for primacy effects and 2 were control pairs for recency effects. None of the control pairs (all nonsynonyms) appeared on the test list. The remaining 20 pairs were sources of items for the test list. They consisted of four sets of 5 pairs each. For the first set, the wrong items (all nonsynonyms) entered into the test list as old items, but the right items were replaced by their synonyms (E_R items). For the second set, the right items (all nonsynonyms) entered into the test list as old items, but the wrong items were replaced by their synonyms (E_W items). For the third and fourth sets, both the right items and the wrong items (all nonsynonyms) entered into the test list as old items. Within-set pairings between right and wrong items were determined by random assignment of words from appropriate segments of the total word pool. In addition to the items described above, the test list contained 10 new items that were unrelated to prior study items. Of these items, 5 were synonyms that served as either E_R or E_W items for other subjects, and 5 were nonsynonyms that served as fillers. Thus, the test list was composed of 50 items: 15 prior right (old), 15 prior wrong (old), 5 E_R, 5 E_W, 5 C and 5 fillers.

A pool of 83 predominantly concrete nouns provided the items for the lists. The core words were 20 pairs of familiar synonymous nouns. The individual members of these pairs had frequency counts (Thorndike and Lorge, 1944) ranging from 14 per million to AA. Free-association values (e.g., Palermo and Jenkins, 1964) of the paired synonyms ranged from 2.7% to 70.6%, with a median of 18.1%; and restricted association values (e.g., Riegel, 1965) ranged from 30% to 79%, with a median of 55%. These values are for associations from the member assigned to the study list (stimulus) to the member assigned to the test list (response). The other 43 words in the pool were nouns that matched the general attributes of the synonyms. The nonsynonyms served the functions described earlier. Associative relatedness between items, other than that between synonyms, was minimized by using published norms when they were available. In other cases, decisions about relatedness were based on the experimenters' judgment.

A given pair of synonyms functioned in three different arrangements: one member of the pair a right item in the study list, the other member an E_R item in the test list; one member of the pair a wrong item in the study list, the other member an E_W item in the test list; one member of the pair absent in the study

list, the other member a C item in the test list. Ten subjects at each practice level received each of the three arrangements a given pair entered into. Eight variations of the lists were employed to assure that all 20 pairs of synonyms occurred in each of the three arrangements. The pairs were rotated across the arrangements in four groups of 5 pairs each, the groups' composition being determined by random assignment. Pairings of right and wrong items on the various study lists were also determined by random assignment, but with the restriction that items, whether right or wrong, that were later to be replaced by their synonyms, were always paired with items that were not so to be replaced. Differences in recognition between the eight variations were subsequently found to be negligible, and the variations were pooled at each practice level for all analyses reported later.

Procedure

The general procedure followed closely that of the earlier studies employing a single-item test (Kausler et al., 1975; Kausler and Yadrick, 1976, 1977). It was again implied to the subjects that study and test phases would follow the format of standard multiple-item recognition learning. That pairs could be expected on the test list was implied by the instructions and by the use of a short practice list of paired nonsynonyms.

After looking over the practice list, the subjects received one, two, or four study trials on the 24-pair study list described above. The study pairs were exposed via a Kodak Carousel slide projector at a 3-sec rate. The intertrial interval where appropriate was 18 sec. Intrapair members on the study list were arranged vertically with the right item underlined. A different random serial order was employed on each trial, but with the control pairs always maintaining their positions for primacy or recency. On each trial, 10 of the experimental right items were exposed on top and 10 on bottom. For the groups with two and four study trials, each word appeared equally often on the top and bottom.

After the designated number of study trials, the subjects were informed that they would be tested on 50 individual items, rather than on pairs. They were told that the test items would include both items that had been right and items that had been wrong, as well as new items that had not been included in the study phase. However, they were not told the number of old items embedded within the test list. It was thought that this information might influence their decisions about the synonyms on the test list. As in the earlier studies, the subjects' decisions about both the oldness and the function of items were recorded on an answer sheet. For each test item, the subject decided initially if it was old or new. If the initial decision was that it was old, the subject then decided if it had functioned as a right item or a wrong item on the study list. The individual test items were projected at a 6-sec rate.

The subjects were run in small groups of 2 or 3 for each variation of the lists at each level of practice. Different random orders of study pairs and test items were employed for each small group.

RESULTS AND DISCUSSION

The proportions of subjects in each group falsely identifying test items as old were determined separately for E_R, E_W, and C items. Mean rates of

false alarms and their standard deviations for the three classes of new test items over the three levels of practice are given in Table 1.

The rates of false alarms were analyzed by a 3 × 3 mixed analysis of variance, with study trials (one, two, or four) as the between-group variable and item type (E_R, E_W, or C) as the within-group variable. This analysis, as well as the analyses of simple effects reported later, was conducted on arcsin transformations of individual subjects' rates of false alarms. However, the results of comparable analyses of the same data without transformations closely approximated those found with the transformations. Rejections of null hypotheses were made with p set at .05.

The main effect of study trials was significant [$F(2, 117) = 20.87$, $MS_e = .055$], as was the main effect of item type [$F(2, 234) = 6.39$, $MS_e = .023$]. The main effect of study trials indicates that subjects became increasingly more accurate in identifying the newness of test items, regardless of type, as the number of opportunities for study increased. A Newman-Keuls test revealed that the overall rate of false alarms was significantly greater after one study trial than after either two or four study trials. However, the difference in rates between the groups with two and four study trials did not attain significance. From Table 1 it may be seen that the effect of item type reflects an overall greater rate of false alarms to the experimental items, E_R and E_W, than to the control items. This disparity between experimental and control items defines, of course, the standard false-recognition effect. Interpretation of the present effect is complicated, however, by the fact that the interaction of study trials by item type was also significant [$F(4, 234) = 2.52$]. From Table 1 it is apparent that this interaction reflects a changing pattern of false alarms as the number of study trials increased. Consequently, the simple effects of item type were examined separately for the various levels of study trials.

After one study trial, the effect of item type was significant [$F(2, 78) = 5.20$, $MS_e = .025$]. Contrasts between means (Newman-Keuls test) yielded a significantly higher rate of false alarms to E_R items than to both E_W and C items. However, the rate of false alarms to E_W items was not sig-

Table 1. Mean rates of false alarms (and *SD*s) to synonyms of right (E_R) and wrong (E_W) items and to control (C) items

Item type	Study trials: One	Two	Four
E_R	.475(.267)	.240(.241)	.185(.209)
E_W	.355(.210)	.245(.233)	.160(.213)
C	.325(.190)	.140(.158)	.150(.134)

nificantly greater than that to C items. The effect of item type was also significant after two study trials [$F(2, 78) = 4.40, MS_e = .025$]. More important, the rate of false alarms to E_W items, as well as to E_R items, significantly exceeded that to C items, and the rates for E_R and E_W items did not differ significantly. The groups with one and two study trials were undoubtedly responsible for the overall main effect of item type reported earlier. An orthogonal components analysis of this main effect revealed further a significantly greater rate of false alarms to the combined E_R and E_W items than to the C items [$F(1, 234) = 8.96$]. However, the contrast between E_R and E_W items did not reach significance [$F(1, 234) = 3.83$].

On the other hand, no significant effect was found for item type after four study trials [$F(2, 78) < 1, MS_e = .019$]. That is, there was apparently no false-recognition effect for synonyms of either right or wrong items after extended practice on the study list. The rates of false alarms were, in fact, low for all three types of new test items (see Table 1). A decreasing rate of false alarms with an increasing number of study trials for new test items unrelated to prior study items was also found by Kausler and Yadrick (1977). The absence of a false-recognition effect and the presence of an overall reduction in false alarms after multiple study trials may be explained in terms of a model of individual item recognition that is derived from the concepts of signal-detection theory (Kausler et al., 1975). According to the model, the decisions whether individual test items are *old or new* after practice on paired study items are made with respect to a relatively *low* criterion value of frequency units. Items with features carrying frequency values that exceed this criterion are identified as old (i.e., as having been in the study list). Subjects seemingly set a more stringent criterion value after multiple study trials than after a single study trial (Kausler and Yadrick, 1977). Consequently, the proportion of control items exceeding the criterion, and therefore falsely identified as old, is less after multiple trials than after a single trial. More important, since experimental items have only a segment of their features tagged by frequency units, they are also unlikely to have a large proportion exceeding the stringent criterion set after extended practice on the study list. The net effect is a pronounced reduction in the false recognition, on the test, of experimental items, whether related to right or to wrong study items.

The results for the groups with one and two study trials suggest that the depth of featural processing does not differ substantially between right and wrong items. That is, semantic features of both kinds of items are seemingly extracted and tagged with frequency units during practice in standard multiple-item recognition learning. Since wrong items are responded to less often than their right items on each study trial, their

semantic features require more trials than do those of right items if they are to accrue sufficient frequency units to evoke false recognitions as old items. The earlier finding (Kausler et al., 1975) of an increase in false recognition of homophones of right items, but not of homophones of wrong items, probably resulted from the use of only one study trial. Given the semantic processing of both right and wrong items, it seems unlikely that sensory processing would be restricted to right items.

Commonality of sensory and semantic featural processing of right and wrong items does not necessarily mean that there are no qualitative differences in processing the two types of items. Previous studies (Kausler and Yadrick, 1976; 1977) have suggested that cues other than quantitative cues to frequency supplement functional identifications of right and wrong items, although the source of these supplementary cues remains unknown. For example, the evidence in Kausler and Yadrick's (1977) study consisted of progressive increments over study trials in the accuracy of judgments of function of both right and wrong items that were carried over to their test list. The hits correctly identifying right items as right were .588, .690, and .838 after one, two, and four study trials respectively. The false alarms incorrectly identifying wrong items as right were .220, .152, and .116 after one, two, and four study trials respectively. Kausler and Yadrick noted that such increments are unlikely to occur for both types of items if subjects rely solely on frequency to mediate their judgments of function. The use of varying numbers of study trials in the present study permitted a similar analysis of judgments of function for those right and wrong items that were included in the test list. This analysis yielded a pattern nearly identical to that obtained earlier, even though different words entered into the two studies. The hits (and *SD*s) for right items were .602 (.207), .715 (.215), and .793 (.174), and the false alarms for wrong items were .209 (.124), .169 (.143), and .106 (.091), again after one, two, and four study trials respectively. Further analysis and discussion of these results, however, is beyond the scope of this paper.

Finally, of secondary interest in the present study is the effect of item type on the judgments of function (i.e., right or wrong) of those new test items that had been falsely recognized as old. After one study trial, the proportions of "old" items identified as right in prior function were .368, .253, and .235 for E_R, E_W, and C types respectively. The comparable proportions were .354, .225, and .250 after two study trials, and .243, .094, and .065 after four study trials. The fact that individual subjects contributed in varying degrees to these proportions across the three item types made it difficult to test these differences in proportions for statistical significance. Nevertheless, there did seem to be greater bias to judge the E_R

items as right than the other item types at each level of practice. A detailed examination of the mechanism underlying this bias is beyond the scope of this paper. However, it should be noted that this bias is generally consonant with the signal-detection model mentioned earlier. According to the model, the decisions whether individual test items are *right or wrong* are made with respect to a relatively *high* criterion value of frequency units. Items with features carrying frequency values that exceed this criterion are identified as right (i.e., as having served as a right item in the study list). Since E_R items are presumed to have semantic features with greater frequency values than the semantic features of other types of new test items, they should also have a greater proportion of their total number exceeding the high criterion value.

Notes

This research was supported by National Science Foundation Grant BMS75-05007 awarded to the first author. Requests for offprints should be sent to Donald H. Kausler, Department of Psychology, University of Missouri, Columbia, Missouri 65201. It is the third author who is at Columbia College. Received for publication August 10, 1976; revision, October 1, 1976.

References

Craik, F. I. M., and Lockhart, R. S. 1972. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior* 11: 671–684.

Ekstrand, B. R., Wallace, W. P., and Underwood, B. J. 1966. A frequency theory of verbal discrimination learning. *Psychological Review* 73:566–578.

Kausler, D. H., Pavur, E. J., Jr., and Yadrick, R. M. 1975. Single-item recognition following a verbal discrimination study trial. *Memory and Cognition* 3:135–139.

Kausler, D. H., and Yadrick, R. M. 1976. Differential processing of bilingual right and wrong items in multiple-item recognition learning. *Journal of Experimental Psychology: Human Learning and Memory* 2:284–292.

Kausler, D. H., and Yadrick, R. M. 1977. Item identifications following varying study trials on a multiple-item recognition learning task. *Journal of Experimental Psychology: Human Learning and Memory* 3:203–210.

Palermo, D. S., and Jenkins, J. J. 1964. *Word association norms*. Minneapolis: University of Minnesota Press.

Riegel, K. F. 1965. The Michigan restricted association norms. University of Michigan, NIMH Grant MH 07619, report 3.

Thorndike, E. L., and Lorge, I. 1944. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University.

Underwood, B. J. 1965. False recognition produced by implicit verbal responses. *Journal of Experimental Psychology* 70:122–129.

AMERICAN JOURNAL OF PSYCHOLOGY
September 1977, Vol. 90, No. 3, pp. 489–497

Failure to confirm Elmes, Greener, and Wilkinson's finding on the spacing effect

Douglas L. Hintzman and Leonard D. Stern
University of Oregon

An important piece of evidence favoring an 'attention' explanation of the spacing effect was reported by Elmes, Greener, and Wilkinson (1972). They found that an item following a repeated item in a list was remembered better if it followed a massed repetition than if it followed a spaced repetition. A series of experiments was done to establish the generality of that result, Experiments I, II, and III in the series representing successively closer approximations to the original study. All three experiments failed to confirm the earlier finding.

A recent evaluation of accounts of the spacing effect (Hintzman, 1976) concluded that only one of the several hypotheses considered had any solid empirical support. This was the voluntary-attention hypothesis, which holds that when the second presentation of an item (P_2) occurs shortly after the first (P_1), the subject devotes less attention to it than he does if the interval between the presentations (P_1-P_2 spacing) is longer.

The voluntary-attention hypothesis consists of two somewhat separable assumptions: first, that encoding is done by a central mechanism of limited capacity which must be shared by successive attention-demanding stimuli; and second, that the allocation of this capacity among stimuli is under voluntary control. The subject presumably chooses (for reasons that are not specified) to devote more attention to delayed than to immediate repetitions, and memory for items whose repetition was massed therefore suffers on a later test of retention.

Support for the second of these assumptions has been more difficult to produce than has support for the first. Hintzman, Summers, Eki, and Moore (1975) attempted to manipulate subjects' voluntary allocation of attention and thus to eliminate or attenuate the spacing effect. Their attempts were unsuccessful, suggesting that the mechanism underlying the spacing effect is not under voluntary control. Evidence favoring the oper-

ation of a limited-capacity mechanism, on the other hand, has come from at least three sources. The first of these is Johnston and Uhl's (1976) report that as the P_1-P_2 interval increased, subjects became slower at detecting auditory signals accompanying P_2. This suggests that subjects have 'excess capacity' to give the secondary task during an immediate P_2, capacity which is not available when P_2 is delayed. The second source is Zimmerman's (1975) finding that with self-paced study, subjects spent less time on P_2 when it immediately followed P_1 than they did when it was delayed. Again, this finding suggests that less processing capacity is allocated to immediate than to delayed repetitions. The third source is Elmes, Greener, and Wilkinson's (1972) finding that 'critical' words (one after each P_2) were recalled better when they followed massed-repetition P_2s than they were when P_1 and P_2 were widely spaced. It was as though the amount of attention given the item following P_2 were inversely related to the amount given P_2 itself — as though 'resting' the central processor during a massed P_2 aided the encoding of the next item in the list.

How convincing is this evidence for the role of a central, limited-capacity mechanism in producing the spacing effect? Unfortunately, there is one criticism to which only the third piece of evidence appears immune. The experiments of both Johnston and Uhl (1976) and Zimmerman (1975) required the subject to do something more than passively study a list of items to be remembered. In the former experiment, there were two tasks to be performed simultaneously; and in the latter, part of the task was to press a button terminating one stimulus and initiating the next. Both the secondary signal-detection task and the self-pacing requirement are 'obtrusive' and may alter the subject's behavior in peculiar ways. Spacing may affect performance on a secondary task, or exposure to P_2 when presentation is self-paced, in ways that are quite unrelated to its effect on retention. (It is interesting to note in this regard that a nonobtrusive measure of attention — number of eye fixations given a repeated picture — apparently shows no effect of spacing; Hintzman et al., 1975, Experiment 3.)

The finding of Elmes et al. (1972) escapes this criticism, for there were no apparent extraneous aspects of the task that might alter subjects' behavior. They simply studied lists of words, presented at a fixed rate, which they then recalled. Recall of repeated words increased with P_1-P_2 spacing, while recall of the critical words following P_2 displayed the opposite trend. The task showing the spacing effect and the task suggesting involvement of a limited-capacity mechanism were the same.

The spacing effect has also been found with various materials other than words and in many tasks other than free recall (Hintzman, 1974,

1976), and it is of some interest, therefore, to learn whether the finding of Elmes et al. can be obtained in experimental situations other than that under which it was originally observed. We have performed several experiments seeking to determine the generality of the phenomenon. None of them, as will be reported here, succeeded in duplicating the earlier finding.

An unpublished pilot study, conducted in our laboratory by Richard A. Block, used vacation slides as stimuli and judged frequency as the dependent variable. As is routine using this task, a sizable spacing effect was found. There was no effect of P_1-P_2 spacing, however, on the judged frequency of pictures that immediately followed P_2. Since the experiment was different in many ways from that done by Elmes et al., the failure to confirm their result is difficult to interpret. The present experiments all conformed to their original study in using words as stimuli; Experiment I measured judged frequency, while Experiments II and III, like the original study, measured free recall.

EXPERIMENT I

METHOD

Subjects

Sixty-two subjects were recruited from around the University of Oregon campus. They were paid for their services and were tested in groups of approximately ten subjects each.

Materials

Stimuli were common three-letter English nouns. They were typed in upper case letters using a bulletin typewriter, photographed, and the negatives mounted in slide frames. Thus, each word appeared light on a dark background. Subjects were shown 160 slides, arranged in two 80-position slide trays. The first 4 and last 4 positions were occupied by filler items, as were 14 other positions distributed throughout the list. The remaining positions were divided into six different blocks (three per tray). Within each block, one word was assigned at random to each of seven *repeated item* conditions. One word was assigned to a frequency of 1; three to a frequency of 2; and three to a frequency of 3. Within either of the latter two categories, the three words were assigned repetitions at spacings of 0, 1, and 5 intervening items. For a word at a frequency of 3, the P_1-P_2 and P_2-P_3 spacings were the same. Randomly paired with each repeated item was a *critical item*, which occurred one time only, immediately following the last occurrence of its repeated-item mate. Thus, there were seven repeated items and seven critical items in each of the six blocks.

There were three different rotations of words among conditions, and approximately equal numbers of subjects were tested using each. Pairs of repeated and critical items were rotated together through the three spacings, with their blocks and frequencies unchanged. A different set of six items at a frequency of 1 was used in each rotation.

A single test form was constructed, on which 108 test words were listed in random order, each followed by the digits 0 through 5. The 108 words included all 84 repeated and critical items, the 12 items at a frequency of 1, and 12 words that had not occurred in the list (a frequency of 0).

Procedure

Subjects were told to study each word, as it was shown, for a later memory test, the nature of which was unspecified. The 160 slides were then shown using a Carousel projector, paced by a timer at a 3-sec rate. There was no break in the series between slide trays. Following presentation of the list, the test forms were distributed and subjects were told to circle for each word the digit (0–5) corresponding to the number of times the word occurred in the list. They were instructed to guess if unsure.

RESULTS

The mean judged frequency for the words at a frequency of 0 was .40. Mean judgments for the other conditions are shown in Figure 1: repeated items in the left panel, and critical items in the right. For repeated items, it can be seen that judged frequency was affected by both frequency and spacing. A linear trend test showed the effect of frequency to be highly significant [$F(1, 61) = 455, p < .001$]. A planned comparison using the coefficients -3, $+1$, and $+2$ for spacings of 0, 1, and 5 respectively was also significant [$F(1, 61) = 48.1, p < .001$; overall $SE = .061$].

The same tests were carried out on critical items. The effect of the

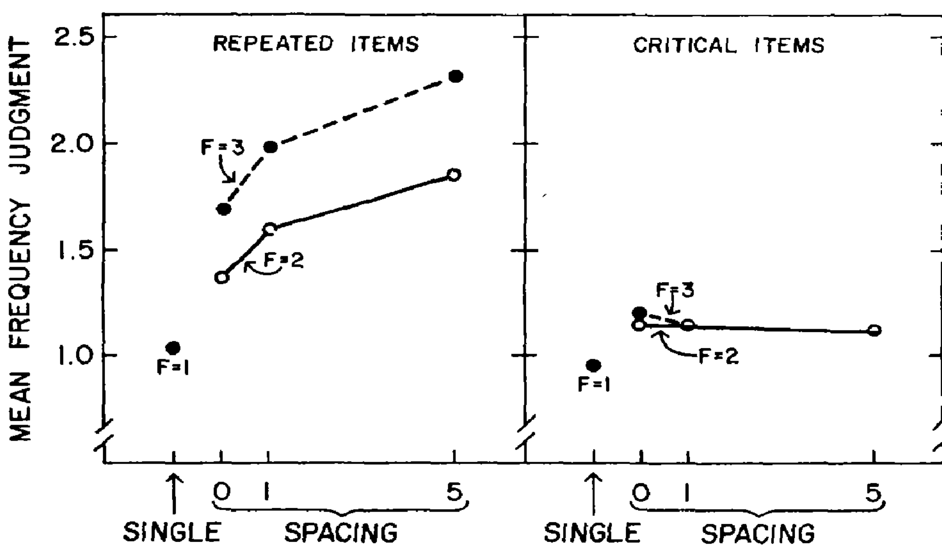


Figure 1. Mean judged frequency, Experiment I

frequency of the repeated items was significant [$F(1, 61) = 13.22, p < .001$]. As Figure 1 shows, this was primarily because the value for the words at a frequency of 1 was low. Since items were not rotated through levels of frequency, this small difference may represent sampling error. The effect of the spacing of the repeated items was not significant [$F(1, 61) = 2.72, p > .05$; overall $SE = .113$].

EXPERIMENT II

Like Block's pilot study, which used pictures, Experiment I showed no effect of the spacing of repeated items on the judged frequency of the critical items. The apparent disagreement with the finding of Elmes et al. (1972) could reflect the use of different dependent variables. Accordingly, Experiment II was done using free recall rather than judged frequency.

METHOD

Subjects

Eighty-six paid subjects, recruited as in the previous experiment, participated. They were tested in groups of approximately ten subjects each.

Materials

Stimuli were the same as those used in Experiment I. Each subject was shown two lists. Each list consisted of 70 slides. The first 10 and last 10 slides, and 4 others throughout the list, were of filler items. Within the middle 50 positions were two blocks of repeated and critical items, representing exactly the same conditions of frequency and spacing as in Experiment I. Thus, there were two lists, and two items per list per condition, or four items per condition for each subject. Three rotations of words among the three spacings were used, as in Experiment I. Approximately equal numbers of subjects were tested on each rotation.

Procedure

Standard instructions for free recall were given, including a warning that some words would be repeated. List 1 was presented at a 3-sec rate, followed by 5 min for written recall. List 2 was then presented and tested in the same way.

RESULTS

Filler words were ignored in the scoring of recall. Figure 2 shows proportions recalled, for both lists combined: repeated items in the left panel, and critical items in the right. For repeated items, the linear trend on frequency was significant [$F(1, 85) = 43.3, p < .001$], as was the planned comparison for spacing [$F(1, 85) = 17.2, p < .001$; overall $SE = .025$].

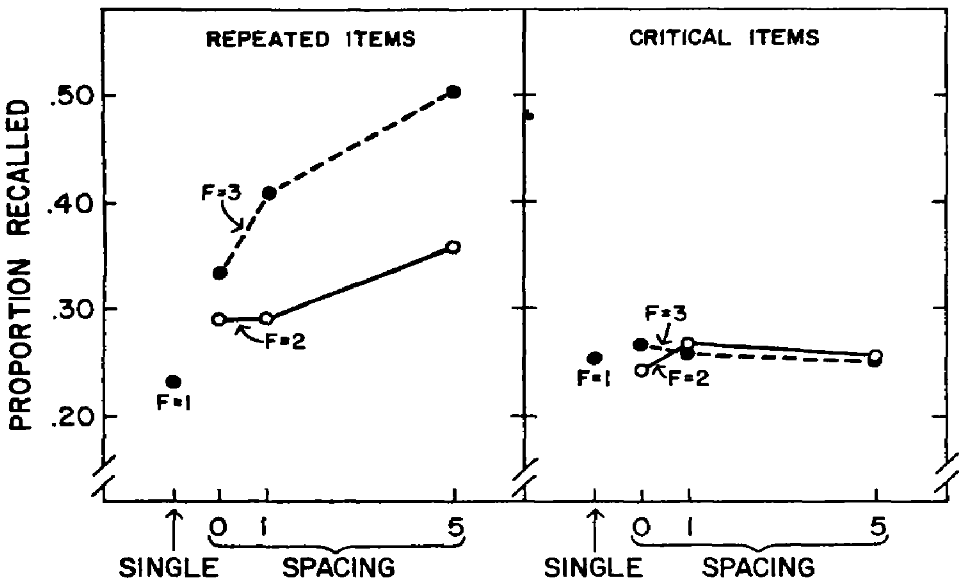


Figure 2. Mean free recall, Experiment II

For critical items, neither frequency nor spacing had a significant effect [$F < 1$; overall $SE = .025$].

EXPERIMENT III

The outcome of Experiment II is clearly at variance with that reported by Elmes et al. (1972). There is, however, one difference between the two studies that might be responsible for the different outcomes. In the study by Elmes et al., each critical item occurred *twice,* following the P_2s of two repeated items of the same condition; in the present Experiment II, by contrast, critical items occurred only once. In Experiment III, this difference was eliminated; each critical item was presented twice.

METHOD

Subjects

Sixty-four volunteers from undergraduate classes at the University of Oregon served as subjects. They were tested in groups of about ten subjects each.

Materials

There were three lists, each consisting of 70 slides, differing in construction from the two lists of Experiment II in only one respect: the same critical item

was used, for a given condition, in both blocks within a list. Thus, there were two repeated items and one critical item per condition per list; or six repeated and three critical items per condition for each subject. As in the previous experiments, repeated and critical items were rotated together through spacing conditions to produce three different rotations. Approximately equal numbers of subjects were tested on each.

Procedure

The experiment was conducted in essentially the same way as Experiment II. The main exception, as noted, was that each critical item was presented twice.

RESULTS

Recall of repeated and critical items is shown in Figure 3. Recall of repeated items was affected by both frequency [$F(1, 63) = 27.1, p < .001$] and spacing [$F(1, 63) = 5.58, p < .05$; overall $SE = .024$]. Recall of critical items was not reliably affected by either frequency [$F < 1$] or spacing [$F(1, 63) = 1.58, p > .10$; overall $SE = .035$]. While the F = 3 curve in the right panel of the figure has roughly the predicted form, the F = 2 curve, which represents conditions most closely approximating those of Elmes et al., does not; and the effect of spacing on both curves combined was not significant.

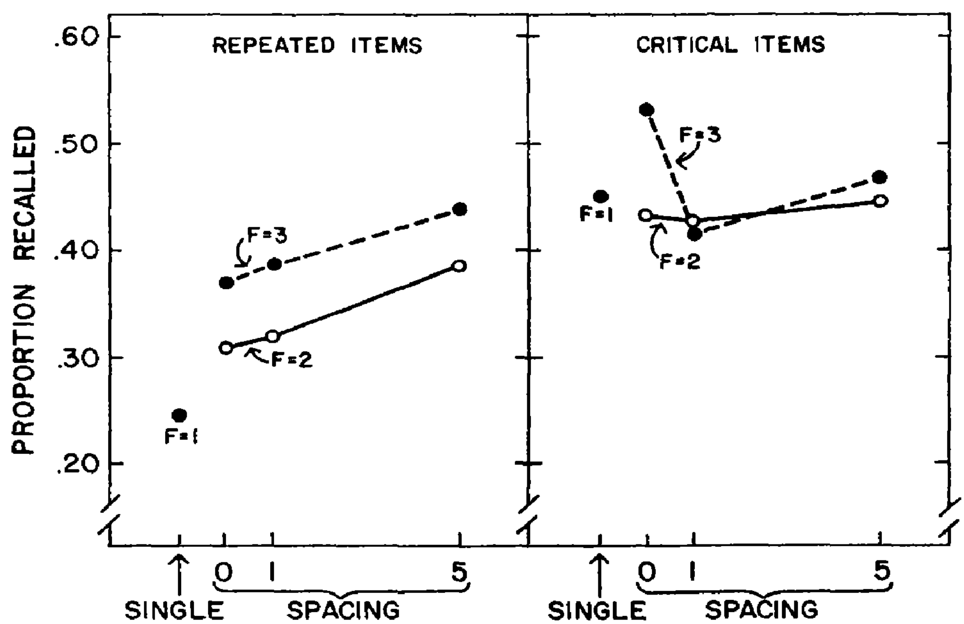


Figure 3. Mean free recall, Experiment III

DISCUSSION

Why did the present experiments fail to obtain the outcome reported by Elmes et al. (1972)? The failure is particularly puzzling in the case of Experiment III, which like their experiments, presented each critical item twice. The answer cannot be that the experiments of Elmes et al. had more power, for the present experiments obtained 372, 344, and 192 observations per condition for critical items, while the two experiments by Elmes et al. obtained 120 and 60. Indeed, their second experiment included only ten subjects, which makes it all the more remarkable that they found a statistically significant effect.

The answer, in fact, may lie in peculiarities of their experimental design. They used very few filler items at the beginnings and ends of their lists, which makes it likely that the strong primacy and recency effects typical of free recall affected their results. Further, while they claim (in their first experiment only) to have used three word orders per list, they do not say how these orders differed from one another. We do not know whether conditions were shuffled systematically among those positions nearest the beginning and end of a list, or even whether words were rotated through conditions. Some words apparently served as a repeated and a critical item simultaneously, in the same list. And since nothing is said about different orders in the description of their second experiment, we must assume that all ten of those subjects were presented with exactly the same lists. Thus, the confounding of spacing with serial position and/or items may have produced the differences among critical items that Elmes et al. attributed to spacing.

All this, of course, is speculation, since we were unable to duplicate the finding of Elmes et al. It may be that some other factor is responsible for the different outcomes of their study and ours. Their lists for free recall were somewhat shorter than those used in the present experiments, for example; and all words in their lists had a frequency of 2, while in the present study, frequencies of 1, 2, and 3 were mixed. It is difficult to see why factors such as these should matter, however. The explanation in terms of unintended confoundings in the study by Elmes et al. is consistent both with the present results and with their published report.

Whatever the reason for the finding of Elmes et al., the generality of that finding appears to be extremely limited. This fact is especially striking when contrasted with the robust nature of the spacing effect itself. Given the present failures to confirm their finding, even under conditions apparently quite similar to those of the original study, it seems unlikely that their finding has anything to do with the cause of the spac-

ing effect. Until more is known about the exact conditions under which their result can be obtained, it is a phenomenon theories of the spacing effect would perhaps do best to ignore.

Notes

Supported in part by Grant GB-40360 from the National Science Foundation. Reprint requests should be addressed to Douglas L. Hintzman, Psychology Department, University of Oregon, Eugene, Oregon 97403. Received for publication October 1, 1976.

References

Elmes, D. G., Greener, W. I., and Wilkinson, W. C. 1972. Free recall of items presented after massed- and distributed-practice items. *American Journal of Psychology* 85:237–240.

Hintzman, D. L. 1974. Theoretical implications of the spacing effect. In *Theories in cognitive psychology: The Loyola symposium,* ed. L. Solso. Potomac, Md.: Erlbaum Associates, 1974.

Hintzman, D. L. 1976. Repetition and memory. In *The psychology of learning and motivation,* vol. 10, ed. G. H. Bower. New York: Academic Press.

Hintzman, D. L., Summers, J. J., Eki, N. T., and Moore, M. D. 1975. Voluntary attention and the spacing effect. *Memory and Cognition* 3:576–580.

Johnston, W. A., and Uhl, C. N. 1976. The contributions of encoding effort and variability to the spacing effect on free recall. *Journal of Experimental Psychology: Human Learning and Memory* 2:153–160.

Zimmerman, J. J. 1975. Free recall after self-paced study: A test of the attention explanation of the spacing effect. *American Journal of Psychology* 88: 277–291.

AMERICAN JOURNAL OF PSYCHOLOGY
September 1977, Vol. 90, No. 3, pp. 499–509

Age-related differences in naming latency

John C. Thomas, James L. Fozard, and Nancy C. Waugh
Thomas J. Watson Research, Yorktown Heights, New York; Veterans Administration Outpatient Clinic, Boston; and University of Oxford, England

Older people seem to have difficulty learning new materials, perhaps because it takes them longer to retrieve relevant encoding information from memory. To assess the effects of age on speed of retrieval, 60 healthy males from 25 to 74 were shown pictures of common objects they were to name aloud as quickly as possible. The older subjects took longer to name the pictured objects. This difference was minimized with practice or when the name was cued, but did not interact with word frequency. The pattern of results for healthy older subjects was not similar to that found by Wingfield for brain-damaged subjects.

The purpose of the present experiment was to study age-related differences in the retrieval of information from long-term memory. Most research directed toward the description of age-related differences in the memory processes of adults has measured the ability of individuals to learn and retain novel or meaningless combinations of letters, numerals, or words (Botwinick, 1973). There are many reasons why older people might encounter greater difficulties memorizing such materials than would young people (see Fozard and Thomas, 1975, for a discussion). For example, learning paired associates allows for the use of different strategies or 'control processes' that affect performance (Hulicka and Grossman, 1967; Hultsch, 1971; Thomas and Rubin, 1973).

In contrast, when overlearned information is retrieved from tertiary (Waugh, 1970) or very long-term memory, differences in strategies associated with age should be small. It is known that an adult's vocabulary score on untimed tests does not change markedly with age. Indeed, vocabulary has been found to be the most stable subtest of the Wechsler Adult Intelligence Scale across age (Botwinick, 1967). For healthy older individuals, there seems to be little doubt that memory for words' meanings remains intact.

In order to study age-related differences in the retrieval of well-learned verbal information, a subject can be shown a pictured object he is to name aloud as rapidly as possible (Oldfield and Wingfield, 1965; Wingfield, 1966). Outside the laboratory, there are two ways such a task is typically made easier. One way is to practice by repeatedly retrieving the name. This also reduces the latency to naming it in the laboratory (at least for younger subjects; Wingfield, 1966, and Bartram, 1973). A second way is to be prepared for retrieving names (e.g., through context). To examine the generality of these approaches, subjects in the current experiment were asked to name a set of pictures over a series of trials. In addition, another condition involved cuing the subjects with the names of the pictured objects.

It is also well known that more frequent words are retrieved from memory more quickly than less common ones (Wingfield, 1966; Landauer, 1975). Some have argued that old people who have slower responses and take longer to learn may suffer from subclinical brain damage. Wingfield (1966) found that patients with diagnosed brain damage take somewhat longer to name common objects but much longer to name uncommon objects. To the extent that older subjects take longer to retrieve information from tertiary memory because of brain damage, one would expect that older subjects would be particularly slow in naming uncommon objects and only a little slow in naming common objects. On the other hand, common words would be expected to reach asymptotic associative strength earlier in life than uncommon ones. In this case, one would expect a flatter slope for the frequency effect in older than in younger subjects. To explore these possibilities, as well as to increase the generality of the results, words from several levels of frequency were used.

METHOD

Subjects

Sixty males ranging in age from 25 to 74 served as subjects. They were divided into five age groups of 12 each: 25–35, 36–45, 46–55, 56–65, and 65+. Measures of verbal IQ (Ammons and Ammons, 1962) were available and the means did not differ significantly for the five groups: 113.2, 114.8, 117.3, 119.7, and 112.8 respectively. All subjects were volunteers participating in an interdisciplinary longitudinal study of healthy aging, the Normative Aging Study (Bell, Rose, and Damon, 1972) sponsored by the VA Outpatient Clinic in Boston. At the time of their entry into the study, all the subjects had been screened on a variety of health criteria that were uniform across ages. The subjects, particularly in the older groups, were therefore generally above average in health compared to an unselected group of their peers. The most frequent occupations of the subjects were insurance worker, policeman, and skilled tradesman.

Apparatus

Slides were presented with a Kodak Carousel slide projector (model AV900). Onset of the stimulus was controlled with a Gerbrands shutter. The subject's vocal responses triggered a Hunter voice key (model 3205), which caused the shutter to close after 100 msec. The slides depicting objects were full-color detailed drawings and, when projected, subtended a visual angle of about 10 deg in width. The pictures were selected from the group created by Shepard (1967) and Fozard (1970). A PDP-8/e computer controlled the projector and shutters and recorded response latencies.

Design

Each subject received the same sequence of stimuli and the same instructions. There were eight blocks of trials that used the same set of stimuli (though in different random orders) and a ninth block that used a novel set of pictured objects. Each block began with 16 *naming* trials, trials in which a subject was shown a picture and asked to name the object as quickly as possible; see the far left of Figure 1. The intertrial interval was 2.7 sec after a correct response and slightly less after an error.

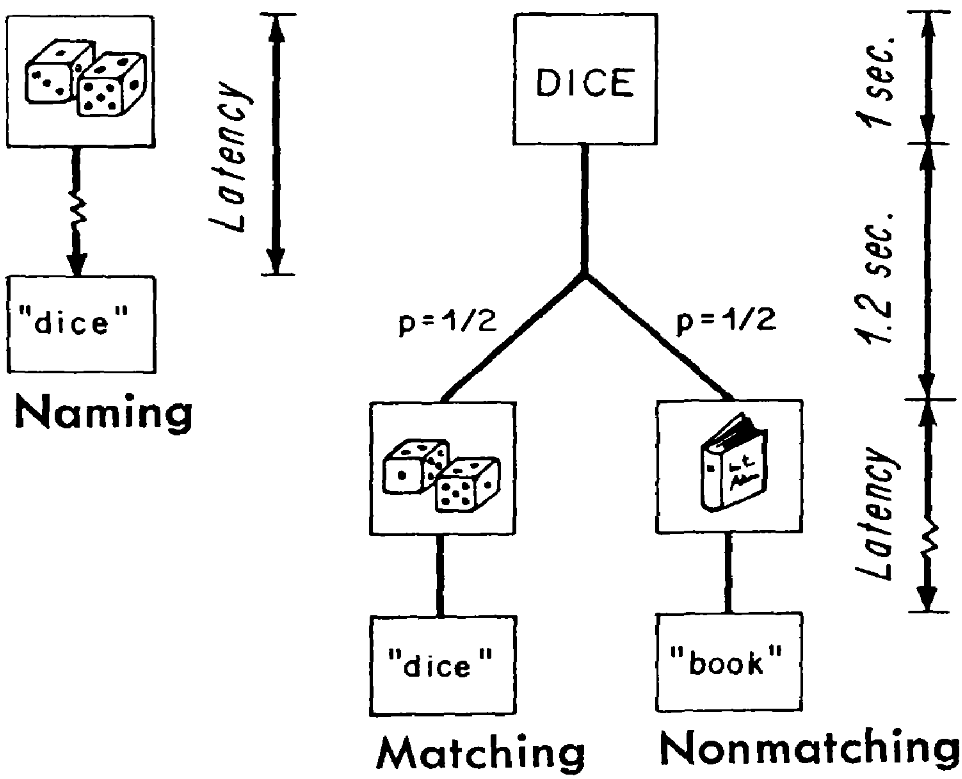


Figure 1. Paradigms for experimental tasks: at the far left, the naming task; at the right, the matching and nonmatching tasks

After the 16 naming trials, another 16 trials followed in which the subject was shown a word prior to each picture; see the right of Figure 1. The word stayed visible for 1 sec, and 1.2 sec after its offset, a picture appeared. The word named the following picture with a probability of .5. Again, the intertrial interval was 2.7 sec after a correct response and somewhat less after an error. The subject's task was always to name the actual pictured object. Trials in which the word shown named the picture that followed it will be referred to as *matching* trials, and those trials in which the word and picture signified different objects will be referred to as *nonmatching* trials.

After a random sequence consisting of 8 matching and 8 nonmatching trials, the next block of trials began. After four blocks, the subject took a brief break while the experimenter changed slide trays. After four more blocks of 32 responses each, the subject was given a ninth block of 16 naming trials followed by 16 matching/nonmatching trials, each with new pictures. At the beginning of the experiment, the subject had been informed that the first eight blocks would use the same pictures though in varying sequence and he was reminded before the ninth block that it would consist of different pictures. In each sequence of 16 pictures, there were two each named by words from eight Thorndike-Lorge categories: AA, A, 40–49, 30–39, 20–29, 10–19, 2–9, and < 1.

Subjects were tested individually. The general purpose of the experiments was first explained to the subject and after his informed consent was obtained, he was conducted to a soundproof room where testing was conducted by a female experimenter. The experimenter monitored the subject's responses and recorded any errors or occasions when the subject accidentally triggered the voice key. If a subject persisted in an error naming a given object for all eight trials across blocks, his response was counted as an error only the first time it occurred. (This happened rarely but would seem consistent with the interpretation that the subject had a different word for the picture than the experimenters.)

RESULTS

Naming task

The mean latencies to correctly naming the pictures for various age groups are shown in Figure 2 for trials across blocks 1 through 8. The data in this figure are averaged over frequency categories. The effects of age were statistically significant [$F(4, 39) = 16.65, p < .001$], as were the effects of block [$F(7, 273) = 54.02, p < .001$]. Although Figure 2 suggests that subjects of all ages replicate Wingfield's (1966) conclusion that even a single experience naming an object greatly shortens response latency, the interaction of age by block was nearly significant [$p = .051$]. This reflects generally smaller age differences on later trials with a given object.

Figure 3 shows naming latencies on blocks 1 and 9 combined as a function of frequency for five widely separated frequency categories. Though words were originally chosen on the basis of Thorndike and Lorge's (1944) frequency count, the placement of the points along the abscissa in Figure 3 is based on the subsequently published norms of Carroll, Davies, and

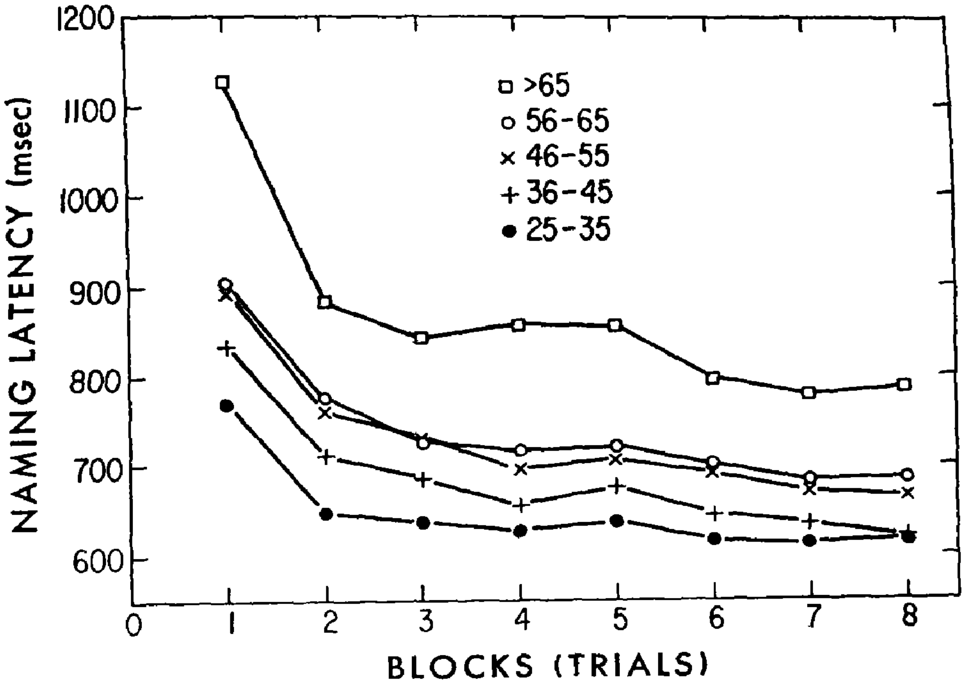


Figure 2. Mean naming latency for various age groups as a function of blocks

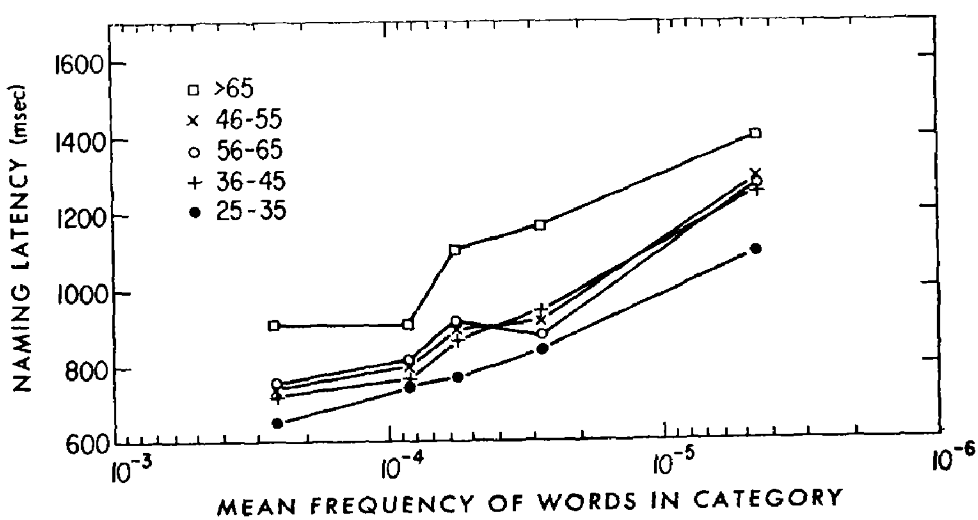


Figure 3. Mean latency for various age groups as a function of frequency, blocks 1 and 9, naming latencies and nonmatching times combined

Richman (1971), norms which allow more exact placement of the AA and A categories. In this graph, and in the analyses of variance, several frequency categories were collapsed to give a more uniform (logarithmic) distribution of frequencies.

Matching/nonmatching tasks

Also of interest was the effect of cuing, or priming, subjects of various ages by presenting the word before the picture. These data, averaged over blocks 1 through 8 are shown in Figure 4. Nonmatching trials took only slightly longer than naming trials. However, matching trials took considerably less time, particularly for the older age groups. In an analysis of variance that included type of task (matching versus naming) as a variable, the main effects of age, block, type of task, and frequency were all significant in the expected direction. The interaction of age by type of task was significant [$F(4, 48) = 4.1, p < .01$], reflecting a larger difference with age in the naming task than in the matching task. The interaction of type of task by block was significant [$F(1, 48) = 24.5, p < .001$], indicating that this difference between the two tasks diminished with practice. The interaction of block by frequency was significant [$F(2, 96) = 4.0, p < .05$], again indicating that the effects of frequency were more pronounced in the early than in the later blocks. Finally, the interaction of type of task by block by frequency was significant [$F(2, 96) = 7.8, p < .001$], reflecting the fact that there was a frequency effect in the naming task that diminished over trials, but no hint of a frequency effect in the matching task.

These results are all consistent with the assumption that the matching task minimized the difficulty of the retrieval involved in naming a pictured object. The results also suggest that the locus of the frequency effect in naming objects is not primarily in the perceptual or motor aspects of the task but in the difficulty of retrieving the names. Apparently, priming a subject eliminates differences in the difficulty of retrieval due to word frequency. On the other hand, part of the difference with age in naming objects may be due to differences in retrieval speed, since reducing the difficulty of the retrieval task (by the priming) reduced the size of the difference.

Individual differences

A variety of additional measures were available for these subjects. Performance on a wide variety of tests of memory and reaction time could be

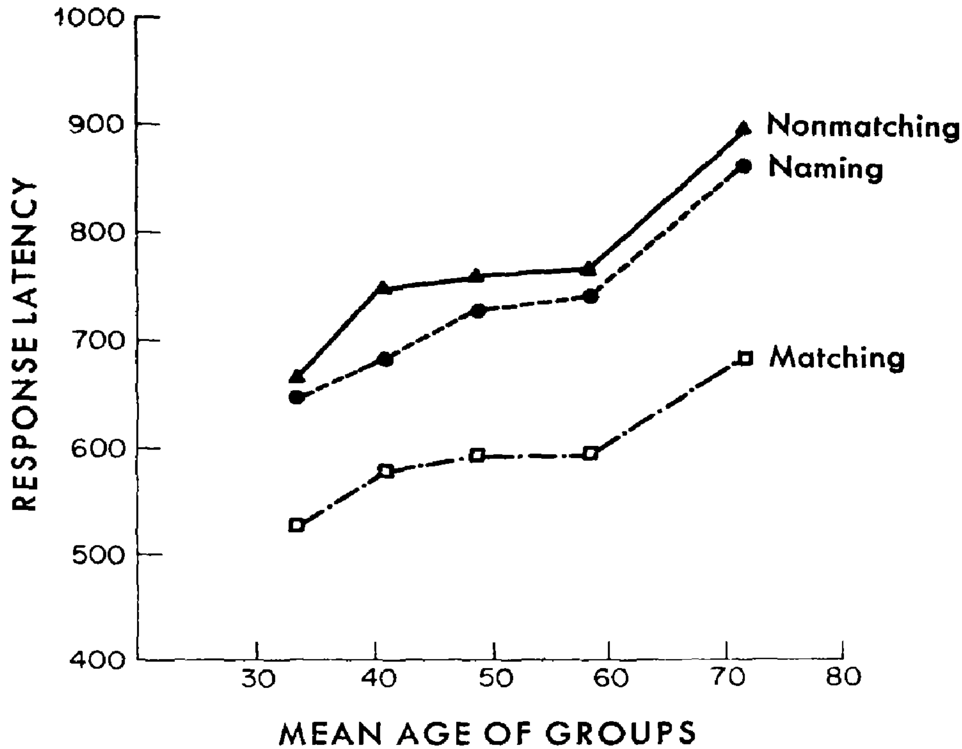


Figure 4. Mean response latency for various age groups for matching, nonmatching, and naming tasks, averaged over the first eight blocks

well predicted on the basis of three or four factors. One of these was an *age/speed* factor. The correlation between these factor scores and the matching times for the subjects was .717 but was only .468 for the naming latencies. On the other hand, the scores for a second factor, *memory,* correlated .572 with naming latencies but only .209 with matching times. This second factor had loadings not only on age but on IQ and on the time necessary to retrieve newly learned paired-associates.

A multiple-regression analysis was run to determine the extent to which individual differences in naming latencies on the first trial with a given object could be predicted. Age alone predicted 40.7% of the variance in mean latency. Verbal IQ (Ammons and Ammons, 1962) added another 7.1% to variance accounted for, and the mean time taken to read common English words accounted for an additional 4.2%. Together these three factors produced a highly significant multiple R of .722. Addition of predictors based on other task performances (of particular interest,

performance in the matching task) did not add significantly to this predictability.

Errors

The overall error rate was low (less than 1% in all age groups except the oldest group, whose overall error rate was 2.8%). The majority of errors occurred in blocks 1 and 9. The highest error rates (6.8%) were found in the oldest group on these blocks. Errors were grouped into one of the following categories: *perceptual,* in which the subject's answer was the name of an object perceptually but not semantically similar to the presented picture; *acoustic,* in which the subject's answer was acoustically similar to the correct answer but not closely related semantically; *superset,* in which the subject failed to give specific enough a name (e.g., 'building' for 'barn'); *semantic,* in which the subject's answer belonged to the same superset category as the correct answer (e.g., 'lamp' for 'table'); *mixed,* in which the subject's answer was related to the correct answer in more than one of the ways listed above (e.g., 'screwdriver' for 'corkscrew' or 'shoe-boot' for 'boot'); *anticipation,* in which the subject's answer was the word presented before a nonmatching trial; and *not classifiable.*

There were differences with age in the prevalence of these errors. In the youngest four age groups, 76% of the errors were basically semantic confusions, such as giving a superset category for an object or giving another item from the same category. With these first four age groups, only 12% of the errors could be considered primarily perceptual in nature. With the oldest age group, the same kinds of errors were present. However, now 35% of the errors were perceptual and only 4% semantic, indicating the possibility of perceptual difficulties for some older subjects.

DISCUSSION

The effect of age on naming latency was not unexpected. Decreasing the difficulty of naming pictured objects by continued practice or by priming the subject (matching task) decreased the effect of age. Further work is needed to understand exactly what the effect of priming is, not only on the retrieval involved in naming a pictured object but on the perceptual task and the organization of a vocal response. It may be that a subject had to do essentially the same processing on the matching task that he did on the naming task and that priming simply made this processing easier. However, several facts throw doubt on this assumption:

first, a person's age and IQ were much better predictors of his performance on the naming task than was his mean time on the matching task, and second, there was no apparent frequency effect whatever in the subjects' performance on the matching task.

Much of the age-related slowing in the naming of pictured objects is probably due to perceptual or motor differences among the age groups. This notion is supported by the fact that although the matching task reduced the difference between age groups, there was still a significant effect of age even when the difficulty of retrieval of names was thus minimized. In addition, the oldest group made substantially more perceptual errors than did younger subjects. In pilot work with young and old subjects, old subjects were found to require a longer presentation time to correctly identify these pictures (84 msec for subjects 19–26 years old, 115 msec for subjects 56–74 years old).

The age-related difference in the latency to name the objects did not interact with frequency, although one might anticipate such an interaction on a priori grounds, or on the basis of Wingfield's (1966) data for brain-damaged patients. In fact, there was no qualitative or quantitative evidence in the current study to suggest similarities between these healthy older adults and brain-damaged patients.

Landauer's (1975) model of memory search is one of several that are consistent with the relationship observed in the present study between word frequency and naming latency. While the present study was not planned to differentiate among models of memory search, it is useful to illustrate how one current model of information processing could be used to relate age and word frequency to naming latency. Structured models of memory (e.g., Collins and Loftus, 1975) could also be used for such a purpose. Landauer's model assumes that multiple copies of the memory traces of words are randomly stored in a three-dimensional memory space in proportion to their frequency of occurrence. During attempted retrieval, a three-dimensional, spreading search starts from a random point and terminates when the correct word is encountered. Two common assumptions about the effects of age can be readily incorporated into this model: the assumption that there is a decrease in the speed of propagation of the search, or the assumption that there is a loss of contents in some of the locations (due to cell death or malfunctioning) and a continuation of the ability to propagate the search. Either of these would lead to a proportional increase in the time needed for search. Older people would need a constant proportion of increased time for search, as compared to that which younger people would need, at each level of word frequency. However, this model's predicted interaction of age with fre-

quency over the range of frequencies studied would be nearly undetectable.

In contrast, subjects sustaining organic brain damage wherein a proportion of their memory locations have not only lost their contents but also their ability to propagate the search would show a disproportionate increase in time to search for low-frequency words. Such a prediction is consistent with Wingfield's (1966) data on brain-damaged subjects. It should be emphasized that the older subjects in the current experiment were exceptionally healthy. In an unscreened sample of elderly subjects, some individuals might show marked decrements in performance. Another caution stems from the fact that error rates were not identical for different age groups or for different frequency classes. Had error rates been constant, larger age-related differences in naming latency and significant interactions with frequency might have been found.

The current results have implications for age-related learning difficulties. Differences with age in the ability to learn new material may be partly due to differences in the ability to encode such information. (See Hunt, Lunneborg, and Lewis, 1975, for recent evidence that the ability to make a quick conversion from a physical representation to conceptual meaning is characteristic of individuals with high verbal intelligence.)

The fact that apparent differences with age in retrieval times were reduced with practice (priming) on particular words could explain why older subjects appear to have difficulty changing 'set.' The practical implication of these findings is that in training older persons, particular care should be taken to allow time for good codes to be retrieved from tertiary memory and that frequent jumps from one topic to another should be avoided. The long latencies and the errors of younger subjects demonstrate that blocking is not a phenomenon restricted to older subjects.

Notes

This paper derives from work at the Department of Psychiatry Research of Massachusetts General Hospital and the Normative Aging Study of the Boston Veterans Administration Outpatient Clinic. The research was supported in part by NIH Grants HD 05669 and MH 08119 and the Council for Tobacco Research — U.S.A. These data were presented at the 25th Annual Meeting of the Gerontological Society, San Juan, Puerto Rico. Karen Graf and Jane Ross assisted in the collection and analysis of the data. Requests for reprints should be sent to John C. Thomas, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598. The second author is also at Harvard Medical School. Received for publication February 10, 1976; revision, October 25, 1976.

References

Ammons, R. B., and Ammons, C. H. 1962. The Quick Test (QT): Provisional manual. *Psychological Reports* 11:111–161 (monograph supplement I–VII).

Bartram, D. J. 1973. The effects of familiarity and practice on naming pictures of objects. *Memory and Cognition* 1(2):101–105.

Bell, B., Rose, C., and Damon, A. 1972. The normative aging study: an interdisciplinary, longitudinal study of aging. *Aging and Human Development* 3(1):5–17.

Botwinick, J. 1967. *Cognitive processes in maturity and old age.* New York: Springer Publishing.

Botwinick, J. 1973. *Aging and behavior.* New York: Springer Publishing.

Carroll, J. B., Davies, P., and Richman, B. 1971. *The American Heritage word frequency book.* Boston: Houghton Mifflin.

Collins, A., and Loftus, E. 1975. A spreading activation theory of semantic processing. *Psychological Review* 82(6):407–428.

Erber, J. T. 1974. Age differences in recognition memory. *Journal of Gerontology* 29(2):177–181.

Fozard, J. L. 1970. Apparent recency of unrelated pictures and nouns presented in the same sequence. *Journal of Experimental Psychology* 86(2):137–143.

Fozard, J. L., and Thomas, J. C. 1975. The psychology of aging: Some basic findings and psychiatric applications. In *Modern perspectives in the psychiatry of old age,* ed. G. Howells. New York: Brunner/Mazel.

Hulicka, I., and Grossman, J. 1967. Age-group comparisons for the use of mediators in paired-associate learning. *Journal of Gerontology* 22:46–51.

Hultsch, D. F. 1971. Organization and memory in adulthood. *Human Development* 14:26–29.

Hunt, E., Lunneborg, C., and Lewis, J. 1975. What does it mean to be high verbal? *Cognitive Psychology* 7:194–227.

Landauer, T. K. 1975. Memory without organization: Some properties of a model with random storage and undirected retrieval. *Cognitive Psychology* 7(4):495–531.

Oldfield, R. C., and Wingfield, A. 1965. Response latencies in naming objects. *Quarterly Journal of Experimental Psychology* 17(4):273–281.

Schaie, K. W., Labouvie, G. V., and Barrett, T. J. 1973. Selective attrition effects in a fourteen-year study of adult intelligence. *Journal of Gerontology* 28(3):328–334.

Shepard, R. N. 1967. Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior* 6:156–163.

Thomas, J. C., and Rubin, H. 1973. Age and mnemonic techniques. *Gerontologist* 13(3):81. Abstract.

Thorndike, E. L., and Lorge, I. 1944. *The teacher's wordbook of 30,000 words.* New York: Teacher's College, Columbia University. 1944.

Waugh, N. C. 1970. Primary and secondary memory in short-term retention. In *The biology of memory,* ed. K. H. Pribram and D. E. Broadbent. New York: Academic Press.

Wingfield, A. 1966. The identification and naming of objects. Doctoral dissertation, University of Oxford.

AMERICAN JOURNAL OF PSYCHOLOGY
September 1977, Vol. 90, No. 3, pp. 511–516

On the inconstant effects of study instructions

James D. Evans
Lindenwood Colleges

This experiment examined the assumption that additional study time will have a greater long-term memory effect under elaborative-rehearsal instructions than under maintenance-rehearsal instructions regardless of whether study time occurs during or immediately after the presentation of a "unit of study." Using either elaborative study or simple rehearsal, 40 subjects free-recalled each of 24 four-word lists, and were then given an unexpected final free-recall test. Study time was varied both during and after list presentation. The results indicated that the interaction of study instructions and study time is constrained by the locus of additional study time in the task sequence.

During the past decade and a half, 'cognitive' approaches to research on verbal learning and memory have become increasingly popular (see Craik and Lockhart, 1972). A major methodological spin-off of the cognitive trend is the ever-growing practice of manipulating instructions for study as an independent variable — instructing half of the subjects to use a mnemonic device while allowing the remaining subjects to use 'simple rehearsal,' for example. The implicit assumption of this manipulation is that the effects of such study instructions will be constant throughout the task. This simplifying assumption has resulted in the posing of some uncomplicated questions such as 'What type of study (information processing) is sufficient for transfer of information to long-term memory?' Unfortunately, investigations derived from this kind of question have yielded conflicting results (compare Darley and Glass, 1975, with Modigliani and Seamon, 1974). The present experiment began as an attempt to investigate the straightforward question mentioned above, and it ended as a demonstration of the fact that the influence of a fixed set of study instructions on *long*-term recall may vary as a function of the locus of the time available to study a list for prior *short*-term recall.

In the original design of the experiment, type of study instructions — for simple rehearsal versus elaborative study — was to be combined orthogonally with two rates of presentation of the items. The ostensible task was the short-term free recall of each of several four-word lists, and the

test of long-term memory consisted of the subjects' unexpectedly being asked for the free recall of all words at the end of the experiment. This design permitted an assessment of the effects on long-term memory of type of study as a function of time for study (i.e., presentation rate). Certain multistore models of memory (e.g., Atkinson and Shiffrin, 1968) predict a significant effect of presentation rate under both types of study, whereas more recent models (Craik and Lockhart, 1972) and data (e.g., Modigliani and Seamon, 1974) suggest that the effects of simple rehearsal, defined as rote cycling of items in short-term memory, might be insensitive to increased time for study, and further, that only an elaborative type of processing will enhance long-term storage.

As an afterthought, the original design of the present experiment was extended to include variation in the amount of time available for study *after* (delay interval), as well as during (presentation rate), the presentation of each four words, to investigate the possibility that the arrangement of study time relative to item presentation might be important.

METHOD

Subjects

The subjects were 40 introductory psychology students, who received extr course-credit points for participating in the experiment.

Design

The design was a 2 × 2 × 2 mixed factorial, in which study instructions (sim ple rehearsal or image formation) was a between-subjects variable, and preser tation rate (2 or 4 sec per word) and postlist delay interval (0 or 8 sec) wei within-subjects variables.

Lists

The 24 four-word lists were randomly constructed from a pool of 640 nour with A and AA ratings in the Thorndike-Lorge count. Both the presentation ra and the delay interval after each list's presentation were varied within the s quence of lists in the following fashion. Half of the lists were presented at a ra of 2 sec per word and half at 4 sec per word. Six of the 12 lists presented at t 2-sec rate and 6 of the 12 lists given at the 4-sec rate were followed by an 8-s unfilled delay interval before they were to be recalled. Each of the remaini 12 lists was to be recalled immediately after its presentation (0-sec delay int val). Ten sec were allowed for this initial free recall of each list, and the peri for recall was always signaled by the word "Go." Each period for initial reca except of the last list, was immediately followed by the word "Ready," whi preceded the first word of the next list by 2 sec.

Two 'forms' of the 24 lists were prepared. For each form, cards with t stimuli printed on them were haphazardly shuffled such that a given word's p sentation rate and postlist delay interval in either set of lists was a matter

chance. So that any effect of position would be unsystematically distributed across all four combinations of presentation rate and delay interval, the positions of lists within each form of the sequence of lists were determined through a 'block randomization' procedure. That is, the lists on each form were divided, using a table of random numbers, into six blocks of four lists each, with the constraint that all combinations of rate and delay must appear in every block. Within any block, the positions of the respective combinations were also determined on the basis of a table of random numbers. The lists were then recorded on magnetic tape for auditory presentation. (This block randomization method of balancing for position was preferred over a complete counterbalancing because the former technique minimizes the tendency of subjects to anticipate regular patterns of types of lists and adopt sequential processing strategies.)

Procedure

All subjects were told the number of lists and the lengths of postlist delay interval to expect but were not told of the final test of free recall. The experiment was described to them as a "study of short-term memory." Twenty of the subjects were instructed to simply "rehearse" (i.e., repeat aloud in a cyclic fashion) the items in each list, beginning with the first word presented, until they received the signal for free recall of the four items. The other 20 subjects were told to "process" each list by forming interactive visual images of the four items until the signal for free recall was given. Assignment to these two groups was accomplished on a random basis, and half of the subjects in each group were randomly assigned to each form of the sequence of lists.

All subjects in both groups were given four practice lists. Then they were tested on their *initial* recall (short-term memory) of each of the 24 lists; subjects were tested singly, and the free recall was oral. At the end of the experimental session, each subject was tested on his *final* recall of words from all of the lists; this free recall was written. They were urged to guess when uncertain and were allowed as much time as they wanted to complete this final recall.

RESULTS AND DISCUSSION

The means and variances for initial and final recall, as a function of study instructions, presentation rate, and delay interval are shown in Table 1. Since the eight cells formed by the orthogonal combination of the three independent variables were virtually identical in level of initial recall, and probably reflected a ceiling effect, the data on initial recall were not subjected to a statistical analysis. The data on final recall were analyzed via a 2 $\times$ 2 $\times$ 2 (study instructions $\times$ presentation rate $\times$ delay interval) mixed analysis of variance, in which study instructions was the between-subjects variable and presentation rate and delay interval were within-subjects factors.

As expected from previous research, long-term free recall was better when image formation rather than simple rehearsal was used [$F(1, 38) = 8.45, p < .01$], when words were presented at the 4-sec rate rather than

at the 2-sec rate [$F(1, 38) = 30.29$, $p < .01$], and when 8 sec of additional study time was given after the list was presented [$F(1, 38) = 10.87$, $p < .01$].

Of greater interest are the possible interactions of study time and study instructions. As indicated in the data in Table 1, increasing the amount of study time *during* list presentation appeared to aid long-term recall regardless of whether subjects were instructed to use only simple rehearsal or an elaborative type of processing (i.e., image formation). This observation is supported by the fact that rate of presentation did not interact significantly with study instructions [$F(1, 38) = 1.50$, n.s.], indicating that slowing the presentation rate had the same beneficial effect under both types of study instructions. In contrast, however, providing an additional 8 sec of study time *after* list presentation enhanced long-term recall only when subjects were employing an elaborative form of study [$F(1, 38) = 6.10$, $p < .05$]. A Newman-Keuls analysis clarified the interaction of study instructions and delay interval by showing that the 8-sec delay interval yielded better recall than the 0-sec one at both the 2-sec [critical difference = 1.90, $p = .05$] and the 4-sec presentation rates [critical difference = 1.95, $p < .01$] in the image-formation group, but at neither the 2-sec [critical difference = 1.62, $p > .05$] nor 4-sec presentation rate [critical difference = 1.35, $p > .05$] in the simple-rehearsal group. No other interactions reached statistical significance. Thus, increasing the amount of simple rehearsal during a list's presentation produced results in support of Atkinson and Shiffrin's (1968) model, whereas increasing the amount of simple rehearsal after a list's presentation led to an outcome consistent with Craik and Lockhart's (1972) model.

Table 1. Mean initial and final free recall (and variances) as a function of study instructions, presentation rate, and postlist delay interval

	Presentation rate			
	2 sec per word		4 sec per word	
	0-sec delay interval	8-sec delay interval	0-sec delay interval	8-sec delay interval
Initial recall				
Image formation	23.40(2.04)	23.50(1.84)	23.60(2.56)	23.50(0.68)
Simple rehearsal	22.65(2.77)	23.25(1.25)	23.40(1.41)	23.05(3.32)
Final recall				
Image formation	2.30(4.85)	4.20(10.80)	4.60(6.09)	6.55(7.84)
Simple rehearsal	1.60(3.72)	2.40(5.82)	3.60(7.93)	3 35(5.84)

The data presented here allow three conclusions to be drawn, one methodological and two theoretical. On the methodological side, the results show clearly that the effects of manipulating subjects' strategies for study may interact with the phase of the memory task (even when the subjects think the task will require only short-term processing), and this suggests caution in making the assumption that a fixed set of instructions will elicit the same kind of information processing when different amounts and loci of study time are available for the task.

On the theoretical side, the significant effect of presentation rate in both the simple-rehearsal group and the image-formation group indicates that, up to a point, degree of learning may increase with the amount of time that items are attended to during initial presentation, regardless of the type of memory test anticipated by the subjects and irrespective of study instructions. Thus, effects of presentation rate were evidenced in the simple-rehearsal condition even though those subjects ostensibly did not need to use the extra time afforded by the slower presentation rate to bolster the long-term stability of items. This outcome suggests that even when only minimal processing is required by the task, subjects automatically carry long-term storage further when the presentation rate is slowed. What is more, this automaticity of processing suggests that long-term storage does not necessarily depend on an 'optional' process.

Nonetheless, the interaction of study instructions and study time predicted by Craik and Lockhart (1972) did occur when study time was varied after a list's presentation. Perhaps, then, Craik and Lockhart's distinction between the optional Type I (maintenance) and Type II (trace-elaborating) kinds of processing is applicable only after the subject establishes the 'unit of study' (e.g., list, subset of a list, or single item). If the latter assertion is valid, it helps to reconcile the discrepancy between the results of Modigliani and Seamon (1974), who found no effect on long-term memory of increasing Type I processing, and those of Darley and Glass (1975), who did find a positive effect of prolonging Type I processing. The former researchers manipulated study time following a list's presentation, whereas the latter researchers varied study time during a list's presentation.

Notes

Offprints may be obtained by writing to the author at the Department of Psychology, Lindenwood Colleges, Saint Charles, Missouri 63301. Received for publication August 20, 1976; revision, November 1, 1976.

References

Atkinson, R. C., and Shiffrin, R. M. 1968. Human memory: A proposed syster and its control processes. In *The psychology of learning and motivatior* vol. 2, ed. K. W. Spence and J. T. Spence. New York: Academic Press.

Craik, F. I. M., and Lockhart, R. S. 1972. Levels of processing: A framewor for memory research. *Journal of Verbal Learning and Verbal Behavio* 11:671–684.

Darley, C. F., and Glass, A. L. 1975. Effects of rehearsal and serial list positio on recall. *Journal of Experimental Psychology* 104:453–458.

Modigliani, V., and Seamon, J. G. 1974. Transfer of information from short- t long-term memory. *Journal of Experimental Psychology* 102:768–772.

AMERICAN JOURNAL OF PSYCHOLOGY
September 1977, Vol. 90, No. 3, pp. 517–525

Effects of schedules of reinforcement on hypothesis-refining behavior

Lois M. Freed and William J. Freed
University of Kansas

Hypothesis refining is the combining of two separate rules (conjunctive concept), the combination of which is more accurate than either rule alone. Each of 70 subjects was reinforced independently for either rule alone or their combination under one and only one of these conditions: continuous reinforcement (CRF), fixed ratio 2 (FR2), fixed ratio 4 (FR4), variable ratio 2 (VR2), variable ratio 4 (VR4), counterconditioning (COU), or nonreinforcement (EXT). Conditions FR2 and FR4 produced the greatest number of hypothesis-refining behaviors, and conditions EXT, CRF, and COU produced the least. It is suggested that — within limits — the less frequent the programmed reinforcement, the greater the probability of hypothesis refining.

In response to the controversy on the subject of learning without awareness, Verplanck (1962) set forth a new model for studying awareness in verbal conditioning. To test his hypothesis that awareness is a verbal operant, subject to experimental manipulation as is any other behavior, he developed a paradigm. It is Verplanck's paradigm, not the controversy behind it, that is the subject of this paper.

Verplanck's paradigm used several verbal operants, among them the *notant,* which he defined as "the class of verbal chains which state an order in the environment." In his studies, the notants were hypotheses about the rules to be employed in correctly categorizing a set of cards that displayed objects or figures he called *principles.* The subject was presented two side-by-side piles of these cards, stimulus side down; the cards had been sorted by the experimenter according to some rule or sequence of rules. The subject was told to turn the cards over two at a time and for each pair to state the rule (notant) that distinguished the cards on the left from the cards on the right. The notants in his series of experiments were exemplified by 'cards with borders are on the right' and 'girls are on the left.'

Verplanck (1962) conducted many experiments utilizing this paradigm, one of which uncovered a particular type of verbal operant, later named *hypothesis-refining behavior*. In this experiment, the procedure involved having one notant that had previously been reinforced undergo extinction, while at the same time a new notant became subject to reinforcement. After a number of trials during which the subject had discontinued using the first notant, the appearance of a card with both the first and second principles often led the subject to include the extinguishing notant, together with a new one, as a conjunctive notant.

In another experiment (Verplanck, 1962), when one of the two concurrent principles for stacking the decks was being continuously reinforced, some subjects would remain with that one, unmodified. However, a few other subjects would use the second notant together with the first one, while the first was still under continuous reinforcement. A few subjects referred to this behavior as "refining my hypothesis."

Hypothesis refining may be considered as conceptual behavior. According to Bourne (1966, p. 1), "a concept exists whenever two or more distinguishable objects or events have been grouped or classified together and set apart from other objects on the basis of some common feature or property characteristic of each." We may also assume that conceptual behavior is acquired according to both behavioristic and associationistic theories of concept formation (Bourne, 1966). So, if the production of hypothesis-refining behavior could be controlled by practical means, it could very well prove to be valuable to the subject for developing complex, accurate, and precise concepts. The development of methods to enhance the production of hypothesis-refining behavior may therefore be of importance for educational and therapeutic reasons.

A review of the literature revealed no studies, other than the two described, dealing with this behavior. However, these two experiments suggest that hypothesis-refining behavior may depend upon the schedules by which the individual notants are reinforced. That is, in the first experiment, the first notant was continuously reinforced until the second became subject to continuous reinforcement as the first underwent extinction; in the second experiment, the first notant, but not the second, was continuously reinforced. Both of these experiments produced hypothesis-refining behavior, whereas previous studies using Verplanck's (1962) paradigm but without varying the schedules of reinforcement of the notants did not. Therefore, the present study was designed to test the hypothesis that the initiation and, to some extent, the maintenance of hypothesis-refining behavior depend upon the extrinsic schedules used to reinforce each of two concurrent notants.

METHOD

Subjects

The subjects were 70 students at the University of Kansas. They participated in the experiment to fulfill part of the requirements of their introductory psychology course.

Materials and apparatus

Stimulus materials consisted of a set of 90 slides, projected as 17-by-17-in. displays on a screen 4 ft from the subject. Each slide displayed two 3-by-5-in. index cards placed vertically side by side. Two pictures were glued onto each index card. The pictures had been cut from a Sears catalog, a child's picture dictionary, or a gardening magazine. Each slide was different from all others except that either a flower or a person appeared consistently on one slide; however, no two flowers or people were the same. The time from the onset of one set of stimuli to the onset of the next was fixed at 8 sec. Each of the subjects sat at a table and responded orally, his responses being recorded on tape. After an incorrect response, a red light on an aluminum panel on the table in front of the subject was illuminated for .5 sec. After a correct response, a green light that was just under the red light on the same panel was illuminated for .5 sec and a counter to the right of the two lights on the panel was advanced one number, clicking audibly. The lights and counter were controlled manually by the experimenter from an adjoining observation room.

Design and procedure

The subjects were randomly assigned to six experimental and one control condition. The experimental conditions were groups CRF, FR2, VR2, FR4, VR4, and COU, and the control condition was group EXT. The subjects in each of these seven groups were divided into two subgroups, one subgroup with the two relevant stimuli on the right card and the other with the relevant stimuli on the left card, to counterbalance for position. This created a 2×7 factorial design. Continuous reinforcement (group CRF) and counterconditioning (group COU) were chosen because these treatments had previously been found to produce hypothesis-refining behavior, while the fixed- and variable-rate treatments (FR and VR) were used to represent the basic intermittent schedules. VI and FI treatments were not used because with the fixed 8-sec exposure of each slide, VI and FI would be virtually identical to VR and FR.

All subjects were tested individually for one session, during which each subject was presented a series of slides previously arranged according to a sequence of rules. The instructions were: "On the screen before you will appear a series of slides, each slide showing the backs of two cards, side by side. All the cards on the right differ in a systematic way, that is, in the same way, from all the cards on the left. Your task is to view these cards, two at a time, and for each pair tell me the rule that you think distinguishes all the ones on the right from all the ones on the left. (Only state one rule for each slide.) On the table in front of you is located a box with two lights and a counter. If you make a correct hypothesis, the green light will come on, and the counter will advance one number, indicating you have earned one cent. If your hypothesis is wrong, only a red light will come on. Occasionally, after stating a hypothesis, neither light will

come on. This simply indicates that I'm not going to give you any feedback for that particular hypothesis. During the session it is possible to earn a maximum of 85¢." The instructions were typed and a copy given to each subject. In response to questions by the subjects, the experimenter referred them back to the written instructions.

The general procedure for groups FR2, VR2, FR4, and VR4 was the same for each group except, of course, for the particular schedule being used. The first notant (people on the right, or left) was continuously rewarded by the experimenter for the first five correct statements of that principle. Then, the first notant was put on the designated schedule, and at the same time, the second notant (flowers on the left, or right) became subject to reinforcement, being reinforced on the same schedule as the first notant, including continuous reinforcement of its first five statements by a subject. While the first and second notants were being reinforced, if correct hypothesis-refining behavior (people and flowers on the right and left, or left and right) occurred after the first five reinforced trials on the second notant, it was also subject to reinforcement, by the same schedule as the two individual notants. Neither light came on in the event that hypothesis-refining behavior was emitted when it was not scheduled to be reinforced.

For group CRF, the first notant was continuously reinforced during the *entire session*. After the first notant only had been reinforced five times, the second notant was continuously reinforced also. If hypothesis-refining behavior occurred, it too was continuously reinforced.

For group COU, the first notant was continuously reinforced five times and then the second notant was reinforced while the first notant was being extinguished. Once it occurred, hypothesis-refining behavior was continuously reinforced, the second notant remaining continuously reinforced.

For the control condition, group EXT, the presentation order of the slides was the same as used in the experimental conditions except that the control subjects received no reinforcement for any hypothesis.

For all conditions, the sequence of slides was prearranged, the only manipulation being that of reinforcing or not reinforcing the subject's statement of the correct hypothesis.

RESULTS

A 2 (positions) × 7 (conditions) mixed analysis of variance was performed on the number of hypothesis-refining behaviors the subjects showed. There was a significant main effect of conditions [$F(6, 56)2 = 5.13$, $p < .001$]. No main effect of the position of the relevant pictures at the right or left was seen [$F(1, 56) = 1.99$]. The interaction of condition by position [$F(5, 56) = 1.04$] was also nonsignificant. Twenty-one two-tailed t tests were performed between the means of the various treatments to determine their contributions to the significant effect of conditions. All comparisons were significant except three: groups EXT versus COU, CRF versus EXT, and VR2 versus VR4. A summary of t values is given in Table 1. Note that all t values between groups CRF and EXT, on the

Table 1. The *t* values of multiple comparisons for the main effect of condition

	CRF	EXT	COU	FR2	VR2	FR4	VR4
CRF		.01	2.14*	7.52*****	6.27*****	14.04*****	4.65*****
EXT			1.49	6.79*****	4 95*****	12.64*****	3.80*****
COU				3.50	3.72****	11.89*****	2.60**
FR2					3 36***	4.42*****	1.72****
VR2						9.11*****	.70
FR4							9.16*****

* $p = .01$. ** $p = .02$. *** $p = .005$. **** $p = .002$. ***** $p = .001$.

one hand, and groups FR2, VR2, FR4, and VR4, on the other hand, are highly significant.

Figure 1 shows the mean number of hypothesis-refining behaviors compared with the total number of subjects producing even one hypothesis-refining behavior. In terms of the mean number of hypothesis-refining behaviors emitted, the seven conditions could be ranked as follows, from least to most: groups CRF and EXT (these not being significantly different), group COU, groups VR4 and VR2 (not significantly different), and groups FR2 and FR4.

In general, then, extinction, continuous reinforcement, and counter-conditioning produced the smallest number of hypothesis-refining responses; the variable-ratio reinforcements produced an intermediate number of hypothesis-refining responses, and the fixed-ratio reinforcements produced the greatest number of hypothesis-refining responses.

DISCUSSION

The data suggest that the schedules by which two concurrent notants are reinforced after the first few continuously reinforced responses are important in producing hypothesis-refining behavior within this paradigm. They also suggest a definite order of the effectiveness of the schedules investigated. The degrees of the schedules' effectiveness may be accounted for in terms of the frequency of reinforcement. It semes reasonable that the less often a subject receives reinforcement on one notant, the greater the probability of his trying other notants (or hypotheses), and conversely, that the more often he receives reinforcement on one notant, the less the probability of his trying other notants. In addition, the other notants were not emitted in a random fashion, confirming Miller and Dollard's (1941, p. 28) suggestion that "drive impels the person to make responses to cues in the stimulus situation. Whether these responses will be repeated de-

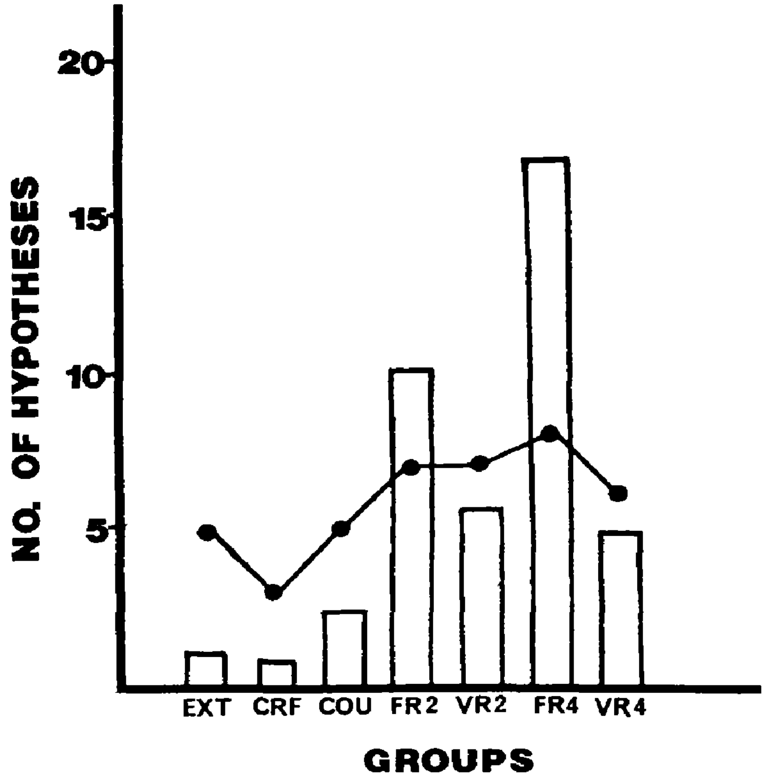


Figure 1. Mean number of hypothesis-refining behaviors for all seven groups (the bar graph), showing the number of people producing the correct hypothesis-refining behavior in each group (the connected data points)

pends on whether or not they are rewarded. . . . As the dominant response is weakened by nonreward, the next response in the hierarchy becomes dominant."

In this experiment, one might consider that a response hierarchy had been established during the first 15 continuous reinforcements, during which each notant and the hypothesis-refining behavior received five continuously reinforced trials. When the frequency of reinforcement decreased on a notant, another notant within the response hierarchy was generally produced until it, too, underwent a decreased frequency of reinforcement, and so on. The notion of the importance of the frequency of reinforcement seems to be consistent with the results obtained for groups CRF, COU, FR2, and FR4, but not for group EXT. Perhaps the latter treatment was ineffective because no response hierarchy had been established, since subjects within this condition were not reinforced for any

response. Groups CRF and COU, which had the greatest probability of reinforcement, showed the fewest hypothesis-refining behaviors. The probability of reinforcement was less for groups FR2 and FR4 than for groups CRF and COU. Group FR2, however, had a higher rate of reinforcement than group FR4, and the total number of hypothesis-refining behaviors was less for group FR2 than for group FR4.

According to this interpretation, the comparison between groups VR2 and VR4 should have been very much like that between groups FR2 and FR4, group VR2 being subject to more reinforcement. However, these two groups emitted essentially similar numbers of hypothesis-refining behaviors. The fact that the groups with variable reinforcement showed significantly less hypothesis-refining behavior may be due to the nature of their schedules. The unpredictability of occurrence of reinforcement may have been a factor in maintaining the individual notants more effectively than the more predictable fixed schedules. An explanation of the lack of a significant difference between groups VR2 and VR4 in hypothesis-refining behavior may also lie within the nature of variable schedules. Morse and Skinner (1957) demonstrated that a few continuous reinforcements within a variable schedule are capable of greatly increasing the rate of responding to the schedule. The fact that each member of groups VR2 and VR4 actually received one or more continuous reinforcements at first may account for the maintained responding with the individual notant(s).

The results of this experiment tend to indicate that hypothesis-refining behavior is effectively produced by particular schedules of reinforcement within this paradigm. In general, the more infrequent the reinforcement (except counterconditioning, the most extreme treatment and one of the least effective, and the variable schedules, for the reasons just mentioned), the more effective the schedule in producing hypothesis-refining behavior.

It would seem that continuously reinforcing a response may suppress possibly more desirable responses and that intermittent reinforcement may produce numerous other, perhaps more desirable behaviors. Such behaviors seem to be facilitated in humans by providing them an opportunity to respond when they do not expect reinforcement for the previous response to be forthcoming.

A wide variety of behaviors not explicitly reinforced have been found to be facilitated by intermittent reinforcement; these behaviors have been termed *adjunctive behaviors* (Falk, 1971). Some adjunctive behaviors that have been demonstrated are schedule-induced drinking (Falk, 1969), wheel running (Levitsky and Collier, 1968), the pica response (Villarreal, 1967 cited by Falk, 1971), aggression (Flory, 1969; Gentry, 1968; Hutchinson, Azrin, and Hunt, 1968; Knutson, 1970), air licking (Mendelson

and Chillag, 1970), and various "interim activities" (Staddon and Simmelhag, 1971). The production of hypothesis-refining behavior reported in the present paper was also facilitated by intermittent reinforcement, here of the individual notants. Since hypothesis-refining behavior was itself reinforced, it cannot be determined whether this hypothesis-refining behavior was similar in other ways to the adjunctive behaviors thus far demonstrated. However, Segal (1972, p. 24) referred to adjunctive and similar behaviors as the product of a process termed "emotional induction, where the bond to inducing stimuli often seems particularly weak," and suggested that this process "might be an evolutionary means of offering up more of the organism's genetically given behavior for operant modification." The view that adjunctive behaviors provide behavioral variability of adaptive value is also shared by Staddon and Simmelhag (1971) and Falk (1971). If hypothesis-refining behavior of the sort produced in the present experiment is indeed adjunctive (at least during its first occurrence for each subject), then it may be an example of how the adjunctive process can indeed provide a clearly adaptive behavior (hypothesis refining) for operant modification during periods of low, but nonzero, reinforcement density.

Notes

This research was supported in part by a research grant from the University of Kansas. The authors thank Dr. Donald Baer for his valuable suggestions, comments, and encouragement. Requests for offprints should be sent to Lois M. Freed, 1402 Fillmore Rd., Oxon Hill, Maryland 20022. Received for publication May 5, 1976; revision, October 6, 1976.

References

Bourne, L. E. 1966. *Human conceptual behavior*. Boston: Allyn and Bacon.

Falk, J. L. 1969. Conditions producing psychogenic polydipsia in animals. *Annals of the New York Academy of Science* 157:569–593.

Falk, J. L. 1971. The nature and determinants of adjunctive behavior. *Physiology and Behavior* 6:577–588.

Flory, R. K. 1969. Attack behavior as a function of minimum inter-food interval. *Journal of the Experimental Analysis of Behavior* 12:825–828.

Gentry, W. D. 1968. Fixed-ratio schedule-induced aggression. *Journal of the Experimental Analysis of Behavior* 11:813–817.

Hutchinson, R. R., Azrin, N. H., and Hunt, G. M. 1968. Attack produced by intermittent reinforcement of a concurrent operant response. *Journal of the Experimental Analysis of Behavior* 11:489–495.

Knutson, J. F. 1970. Aggression during the fixed-ratio and extinction components

of a multiple schedule of reinforcement. *Journal of the Experimental Analysis of Behavior* 13:221–231.

Levitsky, D., and Collier, G. 1968. Schedule-induced wheel running. *Physiology and Behavior* 3:571–573.

Mendelson, J., and Chillag, D., 1970. Schedule-induced air licking in rats. *Physiology and Behavior* 5:535–537.

Miller, N. E., and Dollard, J. 1941. *Social learning and imitation*. New Haven, Conn.: Yale University Press.

Morse, W. H., and Skinner, B. F. 1957. A second type of superstition in the pigeon. *American Journal of Psychology* 70:308–311.

Segal, E. F. 1972. Induction and the provenance of operants. In *Reinforcement: Behavior analyses,* ed. R. M. Gilbert and J. R. Millenson. New York: Academic Press.

Staddon, J. E. R., and Simelhag, V. L. 1971. The "superstition" experiment: A reexamination of its implications for the principles of adaptive behavior. *Psychological Review* 78:3–43.

Verplanck, W. S. 1962. Unaware of where's awareness: Some verbal operants — notates, monents, notants. In *Behavior and awareness,* ed. C. W. Eriksen. Durham, N.C.: Duke University Press.

AMERICAN JOURNAL OF PSYCHOLOGY
September 1977, Vol. 90, No. 3, pp. 527–528

A note on 'graphonyms'

Robert Sommer
University of California at Davis

There are certain rare combinations of longhand letters in English that are structurally ambiguous. Examples of a few (perhaps the only) such graphonyms are presented.

Searching for ambiguous figures, I came across the *urn/win* pattern. When written out in longhand, the two words are identical. I didn't know what to do with it at the time or what to call it. It obviously wasn't a homonym, a pair of words such as *see* and *sea* that look different but sound alike. It did resemble in some ways a homograph, a pair of words spelled the same but with different meanings and pronunciations, such as *row* (rhyming with *know*) meaning to propel a boat with oars and *row* (rhyming with *cow*) meaning a disturbance, but the *win/urn* illusion occurs only when the words are written in longhand. It is based on the special characteristics of certain combinations of written letters. The term 'graphonym' seems appropriate for an illusion based on the characteristics of cursive letters. Following this usage, a pun would be an 'audionym,' a single vocal pattern that can be heard in several ways.

I have found seven combinations of letters capable when written of creating graphonyms. These are *bi/lr, bu/lri, by/bij/lrj, ur/wi, urj/wy, u/ii,* and *d/cl,* shown in that order below. Except for *d/cl,* most of these



combinations are exceedingly rare in the English language. One can find only a handful of words containing *lri* or *wy* in sequence, and very few of these contain the same other letters as the complementary root. This yields only a handful of meaningful English words capable of being confused when written in cursive. Other than the *d/cl* family, I have been able to

locate only three true graphonyms, and all of them involve the letter combination *ur/wi*. Apart from plurals and other modifications of the words themselves, the graphonyms are *win/urn, twin/turn,* and *twined/turned.* The etymological roots of *win, twin,* and *twine* are all independent.



The existence of these few graphonyms in the English language contrasts with the more than 1,300 homonyms and 220 homographs (or homophones as they are sometimes called).[1] The ambiguity of cursive English is thereby minimized. The possibility that a handwritten word will be misread because the letters are combined in the wrong way is very small. Such precision seems necessary because of the tremendous variations in handwriting. It is difficult enough to process the different ways letters are formed without having to worry about ambiguous combinations of letters.

Most ambiguous figures involve some degree of fudging and the graphonym is no exception. The classic duck/rabbit drawing doesn't show a very good rabbit or a very good duck, and Boring's wife has most of her features hidden while his mother-in-law is impossibly ugly. The fudging in the graphonym involves the omission of dots over *i* and *j* and in some cases, a slight rounding of the *r* or an incompletely looped *d*. In the brief period in which words are scanned, these omissions or alterations tend to go unnoticed. If the *i*s and *j*s are emphatically dotted, and the *r*s and *d*s properly formed, then *none* of the fixed combinations of letters mentioned earlier remains identical.

Its rarity is, I believe, the true significance of the graphonym. The English language can be written in longhand with confidence that the letter combinations are not structurally ambiguous. If readers should come across other graphonyms, I should be very interested in hearing of them.

Notes

The author's address is Department of Psychology, University of California, Davis, California 95616.

1. These figures are from Harold C. Whitford, *A Dictionary of American Homophones and Homographs* (New York: Teachers College, Columbia University, 1966).

AMERICAN JOURNAL OF PSYCHOLOGY
September 1977, Vol. 90, No. 3, pp. 529-532

Notes from "Pavlov's Wednesdays": Partial reinforcement as a test of mobility

H. D. Kimmel
University of South Florida

Between late 1931 and early 1935, Pavlov used a partial-reinforcement procedure pairing the conditioned stimulus with food reinforcement on only each fourth presentation. Two dogs learned the differentiation between the first three presentations and the fourth, responding only on the latter; two dogs incompletely learned it; one dog did not. The more successful dogs also had greater mobility, by which Pavlov meant the ability to shift back and forth from excitation to inhibition. As a result of this and related work, Pavlov concluded that mobility, like strength and balance, is an independent primary property of the nervous system and a principal determiner of nervous type.

Although American writers usually credit Pavlov with the first use of a partial-reinforcement procedure (e.g., Kimble, 1961), the scope of his work on this problem and its significance in his definition of types of nervous systems cannot be appreciated from the limited treatment of the subject in English (e.g., Pavlov, 1927, pp. 384–386). Numerous references to experiments involving partial reinforcement are made in the Wednesday seminars (Orbeli, 1949), however, beginning in early 1931 and continuing through February, 1935. The purpose of this note is to summarize this work and explain its significance for Pavlov's theory of types.

The study of partial reinforcement mentioned in 1927 (lecture 22) involved a single dog, one not experimentally naive. Conditioned responses had been established in this dog, using continuous reinforcement, before the partial reinforcement began. A new conditioned stimulus was introduced and was reinforced on every second trial. The dog acquired a conditioned response without difficulty. Next, another new stimulus was introduced, with reinforcement on every third trial. A conditioned response was formed even more rapidly, but the dog became excited. Finally, a fourth new stimulus was added, with reinforcement on only each fourth trial.

Even after 240 trials, no conditioned response was established and the dog became drowsy on the stand. Several tests were undertaken in an effort to comprehend this failure, but none resulted in a satisfactory explanation. The matter was apparently left unresolved for the next several years.

During 1931 and thereafter, Pavlov was actively engaged in reconceptualizing his theory of types. In this context, he reviewed the earlier failure. The dog, Pokorni, was of the *weak* type. Pavlov pointed out that the tactual conditioned stimulus that had been reinforced on only every fourth presentation was associated with a "pathological" cortical point. Even with continuous reinforcement, this stimulus did not acquire a conditioned response. Tactual stimuli in other locations, however, could be made effective conditioned stimuli.

On December 2, 1931, Pavlov described a partial-reinforcement experiment with a new dog, Lis. This dog had been identified as a *strong, balanced* type because it ate from the food trough immediately, unlike more timid animals, and because it made its first conditioned response after only 13 continuously reinforced trials (Gryzun, the all-time record holder, needed only 7 reinforcements before making his first conditioned response). Even for a strong dog like Lis, the procedure was difficult. A visual conditioned stimulus (a light) was interpolated among other, previously conditioned stimuli, some of which were weak and some strong and some of which were excitatory and some inhibitory. In the first series of trials with this procedure, Lis formed a conditioned response to the visual stimulus, in spite of the fact that it was reinforced only every fourth time it occurred. But the response later reduced in strength, as did the conditioned responses to the other stimuli. The differential stimulus was disinhibited. At this point, the dog became very upset and frightened, and it declined the food reinforcement. The visual stimulus was removed, and the dog recovered its normal conditioned responses over a three-week period involving nine experimental sessions. When the light was again introduced with partial reinforcement, the dog again broke down.

Pavlov's interpretation focused on the conflict experienced by the dog and the added difficulty of interpolating the light among the various other, conditioned stimuli. Although he described the interpretation as "subjective," he recognized its mechanistic foundation. "While the light is being repeated three times without reinforcement, extinction of the conditioned responses is occurring; on the fourth, reinforced trial, however, a conflict occurs, an acute clash between the inhibition formed during the extinction trials and the food excitation" (Orbeli, 1949, vol. 1, p. 176).

During 1932 and 1933, Lis and other dogs struggled with variations of this experimental task. A number of animals 'solved' the problem when

no other stimuli except the one receiving partial reinforcement were used. And, ultimately, Lis succeeded as well. But what was the 'solution' of the problem? This is best illustrated in the procedure finally employed with Lis. Salivary conditioned responses were established to a metronome and a bell, using continuous reinforcement. A different metronome was presented without reinforcement until it elicited no reaction. A light was then interpolated among the other three stimuli, with food reinforcement on every fourth presentation of the light (and with the other stimuli continuing to be treated as previously). The intertrial intervals preceding the light varied from one to eight minutes. After considerable difficulty, Lis began to differentiate the first three, nonreinforced administrations of the light from the fourth, reinforced one. The differentiation was stepwise: the first light elicited zero or little response, the second light elicited a small response, the third trial elicited a slightly larger response, and the fourth light elicited a full-sized one. This differentiation was so well established in Lis that the dog was upset when a test involving continuous reinforcement of the light was run; the conditioned responses were temporarily reduced and did not recover to full size until four and a half months of this test were conducted (the number of scale units occurring to the four lights were 60, 55, 50, 55). Within two days after the reinstatement of the partial-reinforcement procedure, Lis's responses to the four lights were 0, 8, 0, and 28 scale units. The next day, they were 0, 6, 23, and 37; the day after that, they were 6, 0, 0, and 43.

Once it was clear that this kind of partial reinforcement would foster the development of a differentiation between the three nonreinforced presentations of the conditioned stimulus and the reinforced one, it was also obvious that only animals with highly mobile nervous systems could succeed at it. At the time Pavlov tried reinforcing only the fourth presentation of a conditioned stimulus, he had not yet recognized the concept of mobility as a significant dimension of higher nervous activity. As a result of the experiments on Lis, his brother Zmei (who learned the differentiation even more easily than Lis), Yulya (another strong dog that acquired the differentiation in a stepwise fashion), and Premier (a strong dog that incompletely learned the differentiation), Pavlov added this partial-reinforcement procedure to the set of diagnostic tests for defining mobility. The three tests were the speed of transformation of differentiated conditioned stimuli (i.e., reversal of positive to negative and vice versa); the speed of formation of long delayed reflexes (e.g., three minutes) interpolated among already acquired short delayed ones (e.g., fifteen seconds); and the acquisition of a differentiation of three identical nonreinforced conditioned stimuli from a fourth one that is reinforced. It should be

clear from these tests that mobility refers to the ease with which the animal can shift from excitatory to inhibitory processes and back.

On December 12, 1934, Pavlov observed that "having discovered mobility," the next task was to decide whether to view it as a completely independent characteristic, analogous to excitatory and inhibitory processes, or . . . as the result of the relationship between excitatory and inhibitory processes" (Orbeli, 1949, vol. 2, p. 592). There were data favoring each of these possibilities, but Pavlov appeared to lean toward the former. He was finally convinced, on February 27, 1935, that mobility was an independent and primary property of the nervous system. The behavior of a dog named Tombush (along with Lis and his brother Zmei) convinced him. Tombush showed perfect retention of a differentiation for over two months without practice and prior to the introduction of any additional differential reinforcement. Because of Pavlov's conclusion that mobility was an independent and primary property of the nervous system, mobility joined strength and balance as the three main determiners of type.

Notes

Professor Kimmel currently is engaged in the first English translation of "Pavlov's Wednesdays," the series of seminars given by Pavlov in the 1930s. The present article describes an aspect of Pavlov's views that is important to both historical perspective and contemporary theory. We are pleased to publish this note, as we were its predecessors, in advance of the complete translation. — N.E.S.

References

Kimble, G. A. 1961. *Hilgard and Marquis' "Conditioning and learning."* New York: Appleton-Century-Crofts.

Orbeli, L. A., ed. 1949. *Pavlovskie sredy* [Pavlovian Wednesdays], vol. 1, *1929–1933;* vol. 2, *1934–1935.* Moscow and Leningrad: Akademiya Nauk, USSR, 1949.

Pavlov, I. P. 1927. *Conditioned reflexes.* Oxford: Oxford University Press.

AMERICAN JOURNAL OF PSYCHOLOGY
September 1977, Vol. 90, No. 3, pp. 533–545

Book reviews

Carl P. Duncan
Northwestern University

The Structure of Human Memory

Edited by Charles N. Cofer. San Francisco: W. H. Freeman, 1976. Pp. 213. $10.00.

The Structure of Human Memory is a title that might lead a naive reader to expect to find some kind of 'wiring diagram of the mind' awaiting him inside the book's cover. But the reader will find something quite different. This may be a cause for either disappointment or relief (or perhaps both), for if the contents are considerably less definitive than a wiring diagram, they are also considerably more readable.

The book is a collection of papers first presented in January, 1975, at a symposium held at the annual meeting of the American Association for the Advancement of Science. This symposium on memory was organized by Cofer as retiring chairman of the Section on Psychology, following a tradition established by Arthur Melton in 1962 (who organized a symposium on short-term memory) and Leo Postman in 1967 (a symposium on the interference theory of forgetting). The book brings together eight papers by leading researchers on cognitive memory. Several chapters (particularly those by Kintsch; Meyer and Schvaneveldt; Schank) are primarily reviews of the authors' work over the last several years. The strong influence of ideas from computer science on current psychological models of memory is represented in chapters by Schank, by Winograd, and by Gilmartin, Newell, and Simon. With the addition of chapters by Estes, by Norman and Bobrow, and by Cofer, Chmielewski, and Brockway, the resulting collection presents an interesting picture of the 'state of the art' in the mid-1970s.

What are the issues that concern these writers? In an introductory chapter, Cofer attempts to relate the various papers to each other and to place the work in historical perspective. He distinguishes between the Ebbinghaus and the Bartlett traditions in research on memory, traditions which roughly mark the ends of a continuum of interests ranging from retention of nonsense syllables to the understanding of stories. The interests of the group of writers represented in this volume generally lie closer to the Bartlett tradition.

The chapter that perhaps best outlines the concerns of current research is that by Roger Schank. He and his colleagues have been concerned with a broad range of issues that includes the representation of words' meanings, the parsing of sentences, the making of inferences based on meaning and world knowledge, the

retrieval of facts and the understanding of stories, and the processes of summarizing and paraphrasing. Furthermore, all of these complex problems are seen as being closely intertwined. While everyone would agree that language is intimately bound up with general knowledge of the world, formalizing this vast body of information in memory has often been considered virtually impossible (e.g., J. J. Katz and J. A. Fodor, *Language*, 1963, 39:170–210). But this is what Schank attempts in introducing the notion of "scripts," which are intended to represent knowledge about stereotyped events (e.g., 'eating in a restaurant'). In whatever way one views the particular models Schank proposes, there is no question that he and others in the study of memory have set themselves challenging goals.

There is an increasing concern with large linguistic units in recent work, a concern reflected in the fact that three of the chapters (by Kintsch; Schank; Cofer et al.) directly involve the memory for prose. Mental processing of any type of material is generally seen as inseparable from effects of the context in which the material is presented. For example, the paper by Estes deals with experimental tasks associated with the Ebbinghaus tradition (free recall and paired-associate learning), but the model he proposes hinges on the notion that contextual cues are integrated into the memory traces for list items. The chapter by Norman and Bobrow argues that "conceptually driven" mental processes, based on expectations created by context and prior knowledge, play a major role in many aspects of memory and perception.

Despite what the title of the book might suggest, no monolithic structure of memory is revealed in these chapters. In fact, the writers approach the issue of structure at several logically distinct levels of analysis. A statement about the structure of memory might take many different forms. These include at least the following possibilities: in terms of a model of the neural substrate of memory, a model of the flow of information between memory stores, a model of memory for particular classes of knowledge, or a model of the associative links within and among memory elements. Let us consider a sample of the views presented in these chapters as they relate to these four levels of analysis.

First, none of the writers attempts to propose a neural model of memory; at some level all speak in terms of information-processing systems. The second possibility, an analysis in terms of a model of memory stores, would describe the flow of information between various distinct types of mental locations, such as a short-term and a long-term store. Such a model is presented in the chapter by Gilmartin et al. 'Box models' of this type have been the target of criticism in recent years, particularly by F. I. M. Craik and his colleagues. The computer-simulation work of Gilmartin et al. suggests one of the reasons analyses of this type nevertheless remain popular, that reason being that the alternatives have not yet been formulated with a comparable degree of precision. An interesting feature of the model of Gilmartin et al. is its explicit provision for variations in subjects' strategies for rehearsal and recall.

At the same time, the model of Gilmartin et al. (called SHORT) also suggests why misgivings about such models persist. Novel features of SHORT (p. 18) include postulated "imagery stores" (auditory and visual) interposed between the more familiar sensory stores and short-term memory. It is not clear what motivates this distinction between imagery stores and short-term memory; for example, rehearsal of words in short-term memory has generally been more

or less identified with a form of auditory imagery. Memory stores sometimes seem to multiply too easily.

The third possible type of analysis of the structure of memory involves distinctions between classes of knowledge in memory. An example might be the distinction drawn in the philosophical literature between 'knowing how' and 'knowing that.' Within psychology, Tulving's distinction between episodic and semantic memory has had considerable influence on recent thinking about memory. These terms were introduced as a heuristic division between memory for personal experiences (episodic) and memory for all 'dateless' knowledge (semantic). Semantic memory encompasses knowledge of rules, procedures for problem solving, and linguistic knowledge. The term semantic memory has also been used more specifically to refer to knowledge of word meanings and inference procedures.

The chapter by Meyer and Schvaneveldt is concerned with the structure of semantic memory in the narrower sense. The bulk of the chapter reviews Meyer's initial paper on the verification of quantified sentences (*Cognitive Psychology,* 1970, 1:242–299). The technique used by Meyer and later investigators is to record reaction times to classify simple sentences such as 'All chairs are furniture' as true or false. The pattern of reaction times obtained across different types of sentences serves as the basis for inferences about the organization of concepts in memory. A major problem that has plagued this line of research since its inception is that a host of highly correlated variables — category 'size,' ratings of semantic relatedness, measures of associative production frequency — can be used to predict the relative speed with which observers can evaluate different subject/predicate relations. The correlational nature of these studies has made it difficult to draw firm conclusions about the nature of the underlying process of evaluation. Meyer and Schvaneveldt, however, have chosen to outline a fairly elaborate model of that process on the basis of Meyer's initial study, with minimal attention to later work, some of which directly challenges the generality of Meyer's findings. The interested reader may consult the more balanced discussion of semantic memory recently provided by Edward Smith (in W. K. Estes, ed., *Handbook of Learning and Cognitive Processes,* vol. 5).

A fourth level of analysis of the structure of memory involves the associative links that connect representations in memory. That is, how are pieces of knowledge 'hooked together' in memory? Proposed answers to that question abound in this volume. They include hierarchies (Estes; Meyer and Schvaneveldt), schemata (Norman and Bobrow), conceptualizations, episodes, and scripts (Schank), and textual bases (Kintsch). The writers use these constructs to talk about problems of perception, comprehension, encoding, and retrieval. In various ways, the constructs are offered as the building blocks of the structure of memory, and they vary considerably in clarity and testability; occasionally, they seem glib (e.g., Schank's statement, p. 180, that "memory is simply a morass of episodes"). In his chapter, Estes voices his concern with how these different approaches relate to each other and to more elementary processes. This concern is an important one, and it is likely to loom larger as proposed mechanisms continue to proliferate. In view of this problem, and given the considerable influence of constructs from computer science on psychological models, Winograd's clear discussion of computer memories and their relation to proposals about human memory is particularly helpful.

The Structure of Human Memory can usefully serve as a source of readings for courses on memory at the advanced undergraduate or graduate level. However, since the chapters are generally tutorials rather than reports of new advances in research, it is likely that investigators will generally prefer to consult the original sources rather than the present volume.

Keith J. Holyoak, *University of Michigan*

Handbook of Learning and Cognitive Processes: Vol. 4, Attention and Memory

Edited by W. K. Estes. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1976. Pp. 436. $19.95.

In an earlier volume of this handbook, which is eventually to be a six-volume review of learning and cognition, the editor describes the progression of the series as moving from the organism's adjustment in the environment to the more abstract features of the organism's cognitive processes. In this volume, the topic of cognition is clearly emphasized; but it is noted in the foreword that "no discontinuity is implied" by this shift to cognitive inqury from the discussions of traditional learning theory found in the previous volumes. Nevertheless, current experimental inquiry into learning and cognitive processes does suffer discontinuity, with disparate histories and the different assumptions these histories bring. For example, the chapters in this volume deal with "attention and short-term memory as these concepts have come to be understood during the period dating from the prescient formulations by Broadbent in 1958 and Melton in 1963" (pp. ix–x). The text, therefore, surveys a youthful enterprise, which is acknowledged by references to a literature that spans less than 25 years. No discontinuity may be implied, but it will be evident in the pages of this volume.

Norman Spear shows that studies of animal memory are also bound by a recent history. He indicates that a psychobiological approach to memory has been outdistanced by the activity and thinking generated in the traditional verbal-learning laboratories. Some of the factors holding back progress in the psychology of animal memory are outlined, but at least until the 1950s they all seem to be adequately described as lack of interest. At about that time, a period well marked by many exciting events in the field of experimental psychology, animal researchers were stimulated by Hebb's neurophysiological theory of memory processing. In the following years, animal researchers became preoccupied with memory consolidation and began the investigation into the microstructure of memory, seeking its structural chemical form. Currently, interest in memory retrieval in animals has been greatly stimulated by the 'reminder' or 'cuing' studies that demonstrate a "reactivation" of original learning. Spear reviews the nature of the treatments used to study memory retrieval in animals and discusses at some length the more recent findings and theories of verbal learning and memory. In fact, one tends to forget that the chapter is about animal memory as Spear wades through many selected phenomena of verbal learning; yet, remarkably, he succeeds in bringing the reader back to the main topic and establishes plausible parallels in animal research. One of the more interesting phenomena — and likely one of the more relevant to the animal studies — is the problem of internal contextual stimuli and human memory. Specifically, the

effects of drugs on memory retrieval suggest converging operations in the animal and human studies. An important aspect of Spear's review is the discussion of problems of separating the effect of the drugged state itself from a real effect on memory. The general field of memory retrieval in animals has concentrated on four major treatments: warm-up, reinstatement, direct reactivation, and implicit reactivation. Discussion of these treatments is well balanced with a review of the attending problems of interpretation.

The following two chapters, "Methodology in the Study of Human Memory," by Murdock, and "The Concept of Primary Memory," by Craik and Levy, provide a brief introduction to the contemporary study of human memory. Murdock's chapter nicely fits a handbook on learning and memory, as it briefly traces the developments in methodology over the last 15 years, a period, according to Murdock, that has seen rapid changes in the ways we investigate memory. Yet, the chapter is not about methodology per se. It is (and this is its strength) about the issues researchers in the field continue to confront. Admittedly a brief chapter, its author nevertheless provides the necessary references for a fuller discussion of the problems he reviews. Murdock brings his own organization to the area by discussing memory in terms of item information, associative information, serial-order information, and free recall.

Recent thinking about the concept of primary memory has resembled the conversation around a campfire at night when everyone will agree that something is out there in the woods beyond their fire but completely disagree as to its shape, size, or the possible consequences of its being there. As Craik and Levy suggest, most researchers are ready to split memory up into two systems, a short-term and a long-term memory. This chapter reviews ideas about this separation. The history of the thinking about short-term memory reveals the varied and often confusing terminology that has attached itself to this concept. At present, four major views of primary memory are available: as a store separate from the long-term store, as a transient trace (involving acoustic or articulatory features), as a different retrieval cue (the least investigated but not necessarily the least interesting of the views), and as a process (*the* current view).

The reader having been introduced to the controversy, the following chapter by Shiffrin elaborates on one particular view of primary memory. According to Shiffrin, the short-term store is a highly limited system giving rise to various phenomena that define the capacity of human information processing. Shiffrin reviews the many areas in which limitations on capacity have been found (e.g., short-term visual and auditory memory, attention and retrieval processes, etc.) and seeks to account for these limitations by the "relatively small set of limitations in a single system: the active memory system, called short-term store" (p. 213). Juxtaposed with Craik and Levy's introduction to this concept, Shiffrin's chapter is the most theoretically stimulating of the volume. The short-term store is not really separate from the long-term store but represents, in Shiffrin's view, an "activated subset." After the automatic stages of sensory processing, control processes take over in the short-term store to select material for additional processing. An important assumption is that all long-term storage "occurs in combination with the context in short-term store at the moment of storage" (p. 216). Unlike other views of the relationship of the two stores, Shiffrin's 'transfer' of information to a long-term depository is not a movement of information from one location to another but the association of previously unasso-

ciated information already in the long-term store, including, as noted above, the current context provided by the short-term store. Material is encoded (i.e., inactive features in long-term storage are activated) in stages, the sequence of processing being from physical attributes to more abstract, cognitive interpretations. All that is called 'capacity limitation' in the human processor is the result of this activated subset within long-term storage. However, this idea may be a bit too unlimited in *its* purview to allow the limits to be tested.

David LaBerge presents a stage analysis of perceptual learning, discussing feature discovery through early coding processes and, eventually, automatic coding. The latter stage is assumed to reflect the decreasing effect of attentional mechanisms. What changes in perceptual learning is attention. Experiments are reviewed that provide evidence of this automatic, attention-lacking perception.

Another introduction to perceptual processes is Massaro's analysis of auditory perception. It is yet another stage analysis, and it attempts to show how a feature-detection system leads to a percept of sound with its various qualities, including location, as a meaningful element in memory. The distinction is made between auditory and abstract perception that follows a sensory experience. Since the generated abstract memory of the auditory code "corresponds to the primary, immediate, or short-term memory in the prototypical memory model" (p. 313), the theory deserves close scrutiny given Shiffrin's previous assumptions about primary memory.

It seems that any contemporary statement on human memory is incomplete without the most recent output of Wayne Wickelgren's theoretical pen. Preventing any incomplete statement is Wickelgren's discussion of memory dynamics, which "refers to time changes in the level (strength, probability of correct recall or recognition, etc.) of the memory trace being acquired, stored, or retrieved" (p. 322). Wickelgren's discussion is distinguished by his positing a single dynamic trace in associative terms. The classical conceptions of interference (competition, unlearning, blocking) still apply in this view. However, decay as a mechanism of forgetting is not alien to this model. And a 'levels' analysis can also apply, according to Wickelgren, if one assumes that rehearsal may succeed in forming phonetic traces but not "higher" semantic traces. Higher levels of processing, it turns out, despite the previous discussion, are related to greater distinctiveness. The trace may not be remembered longer because it is 'stronger,' but because it is somehow more distinctive.

The final chapter, by Wescourt and Atkinson, deals with a relatively recent technique for investigating memory, namely, the reaction-time procedure for testing retrieval from immediate memory. This procedure emphasizes the retrieval of facts, a process taken by some to be the foundation of cognitive operations underlying thinking and inference. Recognition memory, measured by latency of response to items from small, available sets, is presumably based on both directed and nondirected search. The authors see both self-terminating and exhaustive search as classes of nondirected search; directed search is similar to a direct-access retrieval, perhaps based on item-specific information that serves to identify a memory location. However, it is likely that temporal information (e.g., recency of presentation) serves to influence the search, and a model that also accounts for temporal information will likely prove most effective in explaining recognition memory in the reaction-time task. In as fitting a final line

as there might be for a review of the contemporary field of cognition, the authors state that "there are a bewildering number of alternative models for these results, with no unequivocal basis at present for selecting among them" (p. 409). It is the nature of science to ask questions as well as provide answers, and in this young field, there are many more questions available than answers.

This series of volumes, now two-thirds complete, will make those answers easier to find. The scope of the volumes and their progression, discontinuous or not, provide a view of the current state of the inquiry into learning and cognition that is not to be found elsewhere. The contributions are excellent and provide an informative mix of historical review, summaries of evidence, and stimulating theoretical argument. Even so, the series may be most distinguished by the guide provided in each volume by the editor, who begins the volume with a lengthy survey of what is in store for the reader and where it all fits into this cognitive puzzle. His direction of this series and the direction he provides its readers in these scholarly introductions have ensured a lasting contribution to psychology.

Eugene B. Zechmeister, *Loyola University of Chicago*

Human Stereopsis: A Psychophysical Approach

By W. Lawrence Gulick and Robert B. Lawson. New York: Oxford University Press, 1976. Pp. 292. $15.00.

So often, in science, the antecedents of an important discovery can be traced to a technological innovation that permitted observations from a novel perspective. In visual science, Wheatstone's invention of the stereoscope a little over a hundred years ago proved to be just such a major innovation. With this device Wheatstone was able to demonstrate that transverse disparities between the two retinal images are sufficient to produce an immediate and compelling impression of depth, which we refer to as stereopsis. This discovery, besides dispelling the notion that single vision must involve the stimulation of identical retinal points, provided the impetus for a continuing inquiry into the psychophysics of stereoscopic depth, an inquiry which can be traced from Panum, Helmholtz, and Hering, up to the more recent work of Ogle in the 1950s and 1960s. Throughout this lineage was a common, albeit implicit, emphasis on the role of monocular form in binocular disparity.

That emphasis was challenged with the appearance during the 1960s of a second major technical advance in binocular vision, Julesz' development and refinement of random-element stereograms, generated by computer. The global perception of object depth resulting from these formless monocular half-images forcd a reevaluation of the role of monocular form in stereopsis. Julesz' innovative approach, plus exciting developments in the neurophysiology of stereopsis, sparked a resurgence of interest in human stereopsis, and among those most actively working on this problem have been individuals in the perception laboratories at Dartmouth and the University of Vermont, under the direction of W. L. Gulick and R. B. Lawson. The dandy monograph under review, *Human Stereopsis*, represents the culmination of that work. Although it is not explicitly stated, Gulick and Lawson seem to be aiming at a synthesis of more traditional

contour-oriented approaches to stereopsis with the recent cyclopean approach characteristic of random-element stereoscopy. The book presents the results of ten years of research on the psychophysics of human stereopsis; many of the findings have not been published elsewhere. At the outset, the authors disclaim any attempt at a comprehensive review of stereopsis; instead, their focus is on a handful of related phenomena that bear upon stereoscopic contours and texture gradients.

The book opens with two more or less didactic chapters, one devoted to a terse review of the history of thinking on binocular space perception and the other to an excellent presentation on retinal correspondence. This latter chapter provides the most understandable description I've read of the geometry of the longitudinal horopter; any beginning student of binocular vision who quakes at reading Ogle's chapters on the matter should immediately be directed to Gulick and Lawson. The remaining chapters, however, provide the meat of the book. First, the authors begin by introducing the reader to the concept of global stereopsis à la Julesz. Gulick and Lawson assert, and I entirely agree, that the remarkable feature of random-element stereograms is not so much the resulting sensation of depth but rather the global perception of stereoscopic form that emerges via the disparity between the tiny matrix elements of the two half-images. Indeed, this represents the point of departure for what follows: the authors are concerned with the stimulus parameters that determine the vividness and apparent depth of surfaces defined by stereoscopic contour, where stereoscopic contour refers to a subjective boundary generated by retinal disparity in the absence of abrupt luminance gradients.

Now, it is important to realize that most of the stereograms used by the authors are not of Julesz' variety, for although presented in a dot matrix, the half-images employed by Gulick and Lawson *do* contain monocularly visible density gradients. But it is exactly this property, the definition of homogeneous subareas within the matrix, that allows the authors to assess the influence of factors such as matrix density, luminance, and binocular correlation on the formation of stereoscopic contours. And some of their findings are most interesting. To give just one example, stereoscopic contours can appear across relatively large homogeneous regions of the visual field in which there are no local disparity cues. Moreover, these extrapolated boundaries, defined stereoscopically, are invariably seen as linear, even when neighboring regions of the stereogram imply a curved boundary, such as a circle. In effect, the binocular visual system disobeys the Gestalt principle of continuity. This primacy of linear contour in stereoscopic form would seem to disclose a rather fundamental property of the underlying neural mechanism. Also covered in detail are problems of steorscopic size/distance relationships, directional separation and depth adjacency.

The book is filled with stereograms to permit those readers capable of free fusion to verify many of the conclusions firsthand. In each chapter, the results are prefaced with a clear explanation of the experimental procedures employed, magnitude estimation usually among them. In terms of the significance of the results for models of stereopsis, the reader is pretty much left to his or her own imagination. The authors eschew any attempt at formal theorizing, instead inserting occasional summary statements in the form of theorems.

The authors have given us a nicely written, well-illustrated description of their work on stereopsis, work which has both a traditional and a contemporary flavor.

For students of binocular vision, this book certainly deserves a place on the shelf, perhaps most appropriately between the books by Ogle and by Julesz.

Randolph Blake, *Northwestern University*

Cognition and Reality

By Ulric Neisser. San Francisco: W. H. Freeman, 1976. Pp. 230. $11.00.

In 1967, *Cognitive Psychology* by Ulric Neisser was published. That book was to have a profound influence on the development of cognitive psychology, and it can be categorized as perhaps the most germinal treatise on cognitive psychology produced in the 1960s. Then, in 1976, *Cognition and Reality,* the book under review, was published. The more recent book, says Neisser, "is an attempt to deal with several questions that have seemed increasingly important since my earlier survey of cognitive psychology appeared" (p. xi). Neisser believes that three general issues need to be directly and explicitly confronted: "the conception of human nature that is, or ought to be, implicit in the idea of cognition"; the question of "what is happening in contemporary cognitive psychology, and what are we to think of it"; and "the notion of *information processing,*" which "deserves a closer examination" (p. xi, xii). With respect to all three issues, Neisser finds cognitive psychology sadly deficient. First, he believes that "the actual development of cognitive psychology in the last few years has been disappointedly narrow, focussing inward on the analysis of specific experimental situations rather than outward toward the world beyond the laboratory" (p. xi). Second, "there is no disputing the ingenuity and sophistication of much current research, but there is at least some reason to wonder whether its overall direction is genuinely productive" (p. xi–xii). Third, "the amount and kind of processing that a stimulus is assumed to undergo necessarily depends on related assumptions about the nature of that stimulus" (p. xii), a fact not given sufficient emphasis by cognitive psychologists, Neisser believes.

Finally, Neisser says: "In writing *Cognitive Psychology* a decade ago, I deliberately avoided theorizing about consciousness. It seemed to me that psychology was not ready to tackle the issue, and that any attempt to do so would lead only to philosophically naive and fumbling speculation. Unfortunately, these fears have been realized; many current models of cognition treat consciousness as if it were just a particular stage of processing in a mechanical flow of information. Because I am sure that these models are wrong, it has seemed important to develop an alternative interpretation of the data on which they are based. The reader should be warned that Chapter 5, in which these issues are discussed, presents a personal and unorthodox account of attention phenomena rather than a generally accepted view" (p. xiii). In point of fact, most of *Cognition and Reality* represents a highly personal and also highly unorthodox account, not only of attention but of much of present-day cognitive psychology.

For example, Neisser claims that "contemporary studies of cognitive processes usually use stimulus material that is abstract, discontinuous, and only marginally real. It is almost as if ecological *in*validity were a deliberate feature of the experimental design. Subjects are shown isolated letters, words, occasionally line drawings or pictures, but almost never objects. These stimuli are not brought into view in any normal way. Usually they materialize in a previously blank field, and they often disappear again so soon that the viewer has no chance to

look at them properly. They are drawn as if suspended magically in space, with no background, no depth, and no visible means of support" (p. 34). Neisser then proceeds to quote a lengthy 'method section' from a recent study that used the commonplace piece of hardware, namely, a computer-based CRT display system, currently used in studies of cognitive processing. "Such displays," Neisser argues, "come very close to not existing at all. They last for only a fragment of a second, and lack all temporal coherence with what preceded or what will follow them. They also lack any spatial link with their surroundings, being physically as well as temporally disconnected from the rest of the world. They cannot be touched, cannot be heard, and cannot be glanced at more than once. The subject is isolated, cut off from ordinary environmental support, able to do nothing but initiate and terminate trials that run their magical course whatever he may do. Although the data obtained under such conditions can serve as the basis of much ingenious theorizing, the resulting theories may mislead us. Experimental arrangements that eliminate the continuities of the ordinary environment may provide insights into certain processing mechanisms, but the relevance of these insights to normal perceptual activity is far from clear" (p. 35–36).

That kind of thematic and methodological analysis of current cognitive research permeates *Cognition and Reality* and leaves this reviewer with a considerable sense of 'unreality' as to what Neisser believes constitutes the scientific enterprise. Very briefly, science, whether it be physics or psychology, is concerned with, in its broadest sense, discovering the 'Laws of Nature' that ultimately lead to prediction, control, and understanding. Physics learned early that 'Nature' is very reluctant to yield up such 'Laws' easily and that special methodologies and highly sophisticated hardware were needed to achieve even modest successes. One very simple example should suffice: high-energy accelerators are required to study the basic properties of matter, not naturalistic real-world observation of leaves blowing on the wind. By this single example, I do not wish to leave the impression that psychology can, or even should, emulate physics and need only develop highly sophisticated hardware and associated methodologies, but I am firmly convinced that all science requires methods appropriate to its particular subject matter. The problem is that the major thrust of Neisser's analysis of the deficiencies of present-day cognitive psychology and his plea for naturalistic real-world observation, rather than controlled experimental paradigms, seem to provide little, if any, substitute for the methods currently in use.

To be quite specific, Neisser makes the centerpiece of his model of cognitive psychology 'schema' or 'schemata.' He says: "Because the term has already been used rather widely with a variety of meanings, I will try to define what I mean as explicitly as possible. A schema is that portion of the entire perceptual cycle which is internal to the perceiver, modifiable by experience, and somehow specific to what is being perceived. The schema accepts information as it becomes available at sensory surfaces and is changed by that information; it directs movements and exploratory activities that make more information available, by which it is further modified. From the biological point of view, a schema is a part of the nervous system. It is some active array of physiological structures and processes: not a center in the brain, but an entire system that includes receptors and afferents and feed-forward units and efferents" (p. 54). All of chapter 4 is devoted to explicating this basic concept of schemata, particularly along a de-

velopmental dimension. Thus, it is necessary "to credit even the youngest baby with a certain amount of innate perceptual equipment — not merely with sense organs but with neural schemata to control them" (p. 63).

In treating attention and problems of capacity, Neisser admits his unorthodox approach: "It seems to me that hypotheses like those of Treisman or Deutsch and Deutsch are unnecessary. When perception is treated as something we do rather than as something thrust upon us, no internal mechanisms of selection are required at all. The listener follows a message by picking up the information that specifies it as a separate event, and the information that specifies its content and meaning. . . . Organisms are active: they do some things and leave others undone" (p. 84–85). This analysis seems to me to beg the question. Do we not need to ask why, given that organisms are active, they do action A, and not action B, C, D, or Z?

In subsequent chapters, Neisser deals with topics such as cognitive maps, imaging and remembering, perceiving and using speech. Finally, in the last chapter, "Some Consequences of Cognition," he confronts one of the most crucial issues of psychology and, in fact, of all science: the limitations of prediction and control. Again, his posture is highly unorthodox, if not totally incorrect. He says that "the prediction and control of behavior is not primarily a psychological matter. What would we have to know to predict how a chess master will move his pieces, or his eyes? His moves are based on information he has picked up from the board, so they can only be predicted by someone who has access to the same information. In other words, an aspiring predictor would have to understand the position at least as well as the master does; he would have to be a chessmaster himself! If I play chess against a master he will always win, precisely because he can predict and control my behavior while I cannot do the reverse. To change this situation, I must improve my knowledge of chess, not of psychology!" (p. 182–183). If this analysis of prediction and control is correct, it has far-reaching consequences for psychology and the kinds of models necessary to predict human behavior. At the very least, it would seem to suggest that any behavioral domain would require full analysis not only of the organism but of the knowledge and actions relevant to any given behavior. For example, to predict behavior in scientific endeavors, I must be a better scientist than the scientist whose behavior I am attempting to predict; to predict behavior in language or tennis or golf, I must have better language skills or be better at tennis or golf than the individual whose language or tennis or golf I am attempting to predict. But is it not the case that science, and psychology in particular, is concerned with discovering general functional relationships between independent and dependent variables, and that the discovery of such fundamental relationships will have ultimately as a natural consequence corollaries specifying the whys and wherefores of how particular person A beats particular person B, be it at chess, tennis, golf, or whatever?

In brief, this reviewer found *Cognition and Reality* an enormously unsatisfying book, perhaps in part because of its stark contrast to Neisser's earlier efforts in *Cognitive Psychology*. The earlier work was sharp, crisp, to the point, and highly original and innovative. *Cognition and Reality* seems a pallid, superficial, and somewhat illogical addendum, perhaps better left unwritten.

Mark S. Mayzner, *Loyola University of Chicago*

sms of Learning and Memory

R. Rosenzweig and Edward L. Bennet. Cambridge, Mass.: ¡. Pp. 637. $29.50.

volume. It is ambitious in its scope, covering the major areas eurobiology. It provides a variety of simple or model systems ;ychologist interested in neural mechanisms of memory might 'he case is made for the study of memory in a variety of sys- ates to tissue cultures, and the discussion ranges from the de- ning, to the study of macromolecules, to the study of human ·ological disorders of memory, to the ecological approach to

;idered a somewhat diffuse set of chapters, but the reader who ved, for example, in the chapter on tissue cultures and plas- nd Christian is referred to several other chapters within the particular point in greater detail. I suspect one could establish ıgh the book starting at any particular chapter. The frequent · chapters integrate the book. Perhaps this happened because of a meeting, in June of 1974, at which the participants had ties for interaction, but it seems to me that this volume can or other collections. It has all of the outstanding features that ɜ could be obtained with experts in a variety of fields con- icular topic, and none of the disadvantages of dispersion and The book is, perhaps, not tightly integrated, but it is well ɪnced to provide the reader access to the broad scope of ap- mechanisms of memory.

tents of the volume are worth itemizing to gain some appre- ›pe. The first portion, on human memory, deals particularly syndromes. The second section involves computer-oriented e third section is behavioral, ranging from the ecological ap- classical conditioning, to the approach used in studying de- ls. Then follow chapters on anatomy, focusing primarily on ›f dendritic spines and on chemical events, especially those ;. Synaptogenesis is then discussed, on the grounds that such rovide the synapses that underlie the formation of new learn- approaches are then discussed and their application to plas- bed in certain chapters, particularly the chapter by Seil and ıe by Nelson and Christian. The invertebrate model is con- ıtion to the study of synaptic neurophysiology and the appli- neurophysiology to synaptic plasticity are well documented. ːals with the direct approach to the study of memory in verte- ıts, using either drugs or brain lesions or brain stimulation. , cover the wide range of approaches to the study of neural ıory.

› try to specify the audience for which this book is appropriate. ' extremely useful for an advanced undergraduate seminar or seminar on the neural mechanisms of learning. It is suffi- to provide a good summary of research in this area. It is ıte for a professional in neuroscience who is interested in

applying
the engr
intereste
earlier —
— is par
Rescorla
Levine, c
and Cou
other ap
ume. I r
should b
only to b
volume,
with the

his particular area of specialty to the problem of the physical basis of ım. It think it would be appropriate as well for the psychologist who is l in branching into neurological fields. In this regard, the point made that one can start at any chapter and work his way through the book icularly important. One can start with the chapter on conditioning by and Holland, or the critique of classical learning theory by Garcia and r the chapter on development of learning and memory by Campbell ter; then read in areas of one's own expertise; and then branch out to roaches that are covered in the book. I think this is an excellent vol-:commend it highly and feel that the editors, Rosenzweig and Bennet, congratulated for successfully bringing off this ambitious project. It is : hoped that the National Institute of Education, which supported this vill see fit to provide a second meeting that will result in another book xcellent organization and readability of this one.

Aryeh Routtenberg, *Northwestern University*

Neural Mechanisms of Learning and Memory
Edited by Mark R. Rosenzweig and Edward L. Bennet. Cambridge, Mass.: M.I.T. Press, 1976. Pp. 637. $29.50.

This is a first-rate volume. It is ambitious in its scope, covering the major areas of memory and neurobiology. It provides a variety of simple or model systems with which the psychologist interested in neural mechanisms of memory might not be familiar. The case is made for the study of memory in a variety of systems from vertebrates to tissue cultures, and the discussion ranges from the development of learning, to the study of macromolecules, to the study of human patients with neurological disorders of memory, to the ecological approach to memory.

It might be considered a somewhat diffuse set of chapters, but the reader who finds himself involved, for example, in the chapter on tissue cultures and plasticity by Nelson and Christian is referred to several other chapters within the book that make a particular point in greater detail. I suspect one could establish a simple tree through the book starting at any particular chapter. The frequent references to other chapters integrate the book. Perhaps this happened because the book arose out of a meeting, in June of 1974, at which the participants had unusual opportunities for interaction, but it seems to me that this volume can stand as a model for other collections. It has all of the outstanding features that one might imagine could be obtained with experts in a variety of fields contributing to a particular topic, and none of the disadvantages of dispersion and lack of cohesion. The book is, perhaps, not tightly integrated, but it is well enough cross-referenced to provide the reader access to the broad scope of approaches to neural mechanisms of memory.

The specific contents of the volume are worth itemizing to gain some appreciation of their scope. The first portion, on human memory, deals particularly with neurological syndromes. The second section involves computer-oriented model systems. The third section is behavioral, ranging from the ecological approach, to that of classical conditioning, to the approach used in studying developmental models. Then follow chapters on anatomy, focusing primarily on the development of dendritic spines and on chemical events, especially those of protein synthesis. Synaptogenesis is then discussed, on the grounds that such mechanisms may provide the synapses that underlie the formation of new learning. Tissue-culture approaches are then discussed and their application to plasticity is well described in certain chapters, particularly the chapter by Seil and Leiman and the one by Nelson and Christian. The invertebrate model is considered. Its contribution to the study of synaptic neurophysiology and the application of synaptic neurophysiology to synaptic plasticity are well documented. The final section deals with the direct approach to the study of memory in vertebrates, primarily rats, using either drugs or brain lesions or brain stimulation. These studies, then, cover the wide range of approaches to the study of neural mechanisms of memory.

It is important to try to specify the audience for which this book is appropriate. I think it would be extremely useful for an advanced undergraduate seminar or beginning graduate seminar on the neural mechanisms of learning. It is sufficiently up to date to provide a good summary of research in this area. It is certainly appropriate for a professional in neuroscience who is interested in

applying his particular area of specialty to the problem of the physical basis of the engram. It think it would be appropriate as well for the psychologist who is interested in branching into neurological fields. In this regard, the point made earlier — that one can start at any chapter and work his way through the book — is particularly important. One can start with the chapter on conditioning by Rescorla and Holland, or the critique of classical learning theory by Garcia and Levine, or the chapter on development of learning and memory by Campbell and Coulter; then read in areas of one's own expertise; and then branch out to other approaches that are covered in the book. I think this is an excellent volume. I recommend it highly and feel that the editors, Rosenzweig and Bennet, should be congratulated for successfully bringing off this ambitious project. It is only to be hoped that the National Institute of Education, which supported this volume, will see fit to provide a second meeting that will result in another book with the excellent organization and readability of this one.

Aryeh Routtenberg, *Northwestern University*

ANNOUNCEMENT

Special issue of the journal *Philosophy of Science* to be devoted to the philosophy of psychology

Professor Kenneth F. Schaffner, editor-in-chief of *Philosophy of Science,* wishes to announce a special issue of the journal. The special number, tentatively scheduled for December 1978, will be devoted to issues concerning the *philosophy of psychology*. This theme will be interpreted broadly to include, for example, evaluations of experimental paradigms, issues in psycholinguistics, philosophical discussions of statistical techniques, problems of internal representation, psychoanalytic issues, theoretical treatments of perception, motivation, or behavior, as well as psychologically oriented aspects of epistemological or metaphysical problems. Contributors should follow standard instructions for submissions printed inside the back cover of the June 1977 issue of the journal. Essays must be received no later than May 1, 1978, to permit time for review. Manuscripts should be sent to Professor Kenneth F. Schaffner, 314 Loeffler Building, Department of History and Philosophy of Science, University of Pittsburgh, Pittsburgh, PA 15260.

Freud and Philosophy

An Essay on Interpretation

Paul Ricoeur
translated by Denis Savage

"Paul Ricoeur, a French religious philosopher who has not been analyzed nor practiced analysis, has written an important book, a stimulating tour de force that allows us to envisage both the psychoanalytic body of knowledge and the psychoanalytic movement in a broad perspective within the framework of its links to culture, history, and the evolution of Western intellectual thought. In the course of his book, he makes many observations on what he believes comprises the psychoanalytic ethic and philosophy. . . . I can only feel that Freud would have regarded this critique as a challenge and a tribute."—*Psychoanalytic Quarterly*

Cloth $22.50 Paper $6.95

Yale University Press
New Haven and London

Contents

Fred Attneave and Paul Farrar. The visual world behind the head 549

John C. Schmitt and C. James Scheirer. Task difficulty as a mediator of subject strategy in memory scanning 565

Douglas J. Bobko, H. R. Schiffman, Robert J. Castino, and Wayne Chiappetta. Contextual effects in duration experience 577

John J. Shaughnessy. Long-term retention and the spacing effect in free-recall and frequency judgments 587

Lynn Hasher and Michael Greenberg. Expectancies as a determinant of interference phenomena 599

Lawrence M. Ward, Clare Porac, Stanley Coren, and Joan S. Girgus. The case for misapplied constancy scaling: Depth associations elicited by illusion configurations 609

G. J. Syme and Lesley A. Syme. Spontaneous alternation in mice: A test of the mere-exposure hypothesis 621

Joseph P. Forgas and L. B. Brown. Environmental and behavioral cues in the perception of social encounters: An exploratory study 635

Christine D. Bremer, John B. Pittenger, Rik Warren, and James J. Jenkins. An illusion of auditory saltation similar to the cutaneous "rabbit" 645

Donald A. Walter. Discrete events in word encoding: The locus of elaboration 655

Paul T. P. Wong. Partial reinforcement and the learning of a strategy in the rat 663

Seth N. Greenberg. Competition between memory and perceptual tasks involving physically similar stimuli 675

Peter M. Lisiecki and Terry M. Libkuman. The relative effectiveness of rules 1 and 2 in verbal discrimination learning 689

C. Plug. Number preferences in ratio estimation and constant-sum scaling 699

Hans-Lukas Teuber: 1916-1977 (Karl H. Pribram) 705

Book reviews (Alan S. Brown, Earl A. Alluisi, James V. Wertsch, Linda M. Bartoshuk, Joseph Brožek; books received) 709

Consultants 564

Contributors to Volume 90 737

AMERICAN JOURNAL OF PSYCHOLOGY
December 1977, Vol. 90, No. 4, pp. 549–563

The visual world behind the head

Fred Attneave and Paul Farrar
University of Oregon

After a 5-minute inspection of 7 objects laid out on a shelf, subjects were seated with the objects behind them and answered questions about the locations and orientations of objects by throwing a switch left or right. The "visual image" subjects were told to imagine that the objects were still in front of them and to respond accordingly. The "real space" (RS) subjects were told to respond in terms of the positions of the objects in real space behind them. Thus correct responses (left vs. right) were completely opposite for the 2 groups. A control group responded while facing a curtain concealing the objects. The task was harder, by time and error criteria, for group RS than for the other 2 groups, but not dramatically so. All RS subjects denied using a response-reversal strategy. Some reported translating the objects from back to front and thus responding as to a mirror-image of the array. When this evasion was discouraged, RS subjects typically reported responding in terms of *visual* images located behind them and viewed as if by "eyes in the back of the head." The paradox of a visual image that corresponds to no possible visual input is discussed.

Our internal representation of the world around us is based in part on current sensory input, but in much greater part on past sensory inputs, i.e., upon memory. A remarkable feature of the memory traces is their spatial ordering: objects and landmarks encountered at various times in the past, along with those being perceived at the present moment, seem to coexist in a common, more or less continuous representational space that is presumably homomorphic with physical space. A person who is "well-oriented," who "knows where he is," can point with considerable accuracy and with minimal conscious effort at landmarks that are occluded by walls, trees, or the curvature of the earth; he can even do such pointing in a manner that would be appropriate to an imaginary location and orientation of his body (Attneave, 1972).

What can we say, or find out, about the nature of the system that keeps track of where things outside as well as inside the visual field are located? Since the time of Tolman (1948), this system has often been characterized as a "cognitive map," but if the system is maplike in some respects, it

seems rather unmaplike in others, notably in providing the experience of particular *scenes* consistent with particular locations and orientations of the observer. Thus, Piaget and Inhelder (1956, p. 244) emphasize that the child must develop a "global or comprehensive co-ordination of viewpoints" in order to imagine how a model landscape would look from the other side of the table.

Given a particular viewpoint, real or imaginary, the status of objects outside the angular bounds of the visual field poses a special, neglected problem with which any comprehensive theory must deal. Unlike rabbits, people are unable to see all around their heads at once. Nevertheless, the invisible sector of space behind the head seems phenomenally continuous with the sector encompassed by the visual field in front. Does this representation take the form of visual imagery, or kinesthetic imagery corresponding to orienting or pointing responses, or some form more abstract? Suppose someone looks at a spoon lying to the left of a coffee cup and then turns his back to the objects; now the spoon is on his right and the cup is on his left. If he has a visual image of the objects, and this image corresponds to what he would see if he turned and looked at them (or to what he did see when he was looking at them), he should still most readily, though erroneously, think of the spoon as the left object; thus a task that would require him to locate it on the right should involve some measurable cost similar to that found in typical instances of stimulus-response incompatibility. By this reasoning, visual imagery would appear a somewhat awkward medium for representing the world behind the head; however, spatial imagery more generally seems predominantly visual to most people, and it is further questionable that alternatives such as the motor-kinesthetic would provide the necessary fine structure for representing object *orientation* (e.g., which way the handle of the spoon is pointing) in addition to gross location.

Two experiments have been done employing an elaboration of the spoon-cup paradigm suggested above. The subjects individually studied a horizontal array of 7 real objects. They were then seated with the objects behind them and responded to questions of the form "On which side of the duck is the *shoe?*" or "On which side of the dog is the dog's *tail?*" by throwing a key-switch in the appropriate direction. The time required for each answer was recorded. Subjects in the "visual image" (VI) group were told to picture the objects in front of them, as if they were still looking at them, and respond accordingly. For the "real space" (RS) group, the responses required were exactly the opposite, since they were told to answer in terms of the physical locations of the objects be-

hind them. We wished to determine, by the latency criterion, whether one or the other of these two modes of representation was the more primary or natural. A "control" (C) group responded with the objects still in front of them, but with an interposed curtain. Note that even if the subject's representation had been opposite to the requirements of the task given him, he could have responded *correctly* by a simple reversal at the response level (e.g., "The dog's tail is on the left, so I must throw the key to the right"), but the need to do so would presumably have elevated his reaction time.

It will become evident to the reader that in planning these experiments we entertained some preconceptions that were considerably off target: one outcome of the studies has been the drastic revision of these preconceptions.

EXPERIMENT I

METHOD

Subjects

The subjects were 45 paid student volunteers from the University of Oregon, 15 in each of the 3 groups.

Stimuli

Under a large window extending across one end of the experimental room there was a shelf (actually the top of a cabinet) about 2.3 meters long. Seven objects were placed along this shelf at intervals of about .3m; these were, from left to right: (1) a corncob pipe, oriented with the stem left and the bowl right; (2) a wooden decoy duck, facing left; (3) an oversize pencil, pointing left; (4) an old shoe, with the toe on the right; (5) a small realistic figure of a dog, made of painted plastic, facing left; (6) a large kitchen knife, pointing right; and (7) a brown teapot, with handle on left and spout on right. A strip of white cardboard was placed over the window directly behind the objects, but above it one could look out freely at buildings and trees. This was the only window in the room, and the only major difference between the walls to the left and the right was that the former contained the entrance door at the end distant from the window.

Procedure and apparatus

On entering the experimental room, the subject was seated facing the objects at a distance of about 2.5m and was told to study carefully their various locations and orientations. (With a few exceptions to be noted later, all instructions were similar to those of Experiment II, which will be given in detail.) After 3 minutes he or she was encouraged to walk freely around the room, look at the objects from various angles, and even touch them (without moving them) if he wished to. After a minute of such free movement, he was again seated for a final minute of inspection.

the case of groups VI and RS, the subject was then turned around and with his back to the objects, before a table with a key-switch on it. From this point the instructions diverged: Subjects of group VI were told to visualize the array of objects *in front of them,* as if they were still looking at the things or at a picture of them, and to respond to the questions in terms of the left-right relationships in this image; whereas subjects of group RS were told to imagine the objects (no modality specified) in their real positions on the shelf behind, and respond accordingly. In addition there was a group C that responded *without* turning around, but with an opaque curtain hanging about .8m in front of them and concealing the objects; for these subjects there was no conflict between real space positions and the frontal visual image. All subjects were given sample questions and the correct answers, and it was ensured that they understood the task before proceeding.

The test, identical for all groups, consisted of 44 questions. Thirty of these exhausted the possibilities of the form "On which side of the X is the Y?", excluding those cases in which X was an end object (pipe or teapot), since questions about these could be answered without knowing Y. (The excluded cases were, however, used for sample questions in the instructions and for two initial test questions that were not scored.) The remaining 14 questions (2 per object) concerned orientation and took the form "On which side of the pipe is the pipe's *bowl?*" Note that the correct answer to a question always depended on the final word. Order of the questions was quasi-random, with orientation questions so distributed as to have the same mean list position as location questions.

These 44 questions were read into one of the 2 channels of a tape recorder at intervals of 15 seconds, which left about 10 seconds of silence after each for the subject's response. While recording the list, the experimenter pressed a key as he spoke the accented syllable of the last word of each question, thereby placing a tone on the alternate channel of the tape.

The voice channel of the tape was played back to the subject through a loudspeaker at ceiling level directly in front of him (or behind in the case of the control subjects, who were facing the objects). Each tone on the alternate channel started a Hunter Klockounter, which was stopped by the subject's operation of the key-switch in either direction. Errors (which were infrequent) were recorded as well as reaction times. The apparatus and experimenter were located in a room adjacent to the experimental room, but the door between remained open so that the subject could be observed continuously.

RESULTS

All averaging and testing (in both this and the following experiment) was done on the logarithms of reaction times, which were much more nearly normal in distribution than the raw latencies; thus central tendencies will be reported in the form of geometric means. Results were analyzed separately for the 30 Location questions and the 14 Orientation questions.

Differences between subjects (and hence error terms) were exceedingly great: people treated alike gave mean reaction times (RT) that ranged,

Table 1. Reaction times (geometric means, in msec) from Experiment I

Group	Type of question	
	Location	Orientation
Visual image	1,234	1,280
Real space	1,133	1,133
Control	883	873

fairly typically, from half a second to 3 seconds. Thus differences between the group means shown in Table 1 are not significant at the .05 level: for Location, $F(2, 42) = 2.32$, and for Orientation, $F(2, 42) = 1.87$. Only 1 of the 6 pairwise comparisons between groups yielded a t that would attain the .05 level considered in isolation; this was $t(1, 28) = 2.11$ for VI vs. C on Location questions. We shall return to these data after reporting Experiment II.

The mean number of errors (in 44 responses) was 1.8 for group VI, 3.2 for group RS, and 1.3 for group C. Of all the 45 subjects, 23 made no more than one error. Partitioning the 3 groups thus produced the frequencies shown in Table 2, which indicates that the RS task was more difficult than the others: $\chi^2(2) = 8.72, p < .02$. Having been initially inclined to dismiss errors as negligible, we were somewhat astonished to find that they reliably separated the groups, when RT data did not. Note also that there is little resemblance between the pattern of errors in Table 2 and the (nonsignificant) patterns of RTs in Table 1. A further breakdown of errors by Location and Orientation does not suggest any interaction effect.

Each subject, after completing the task, was questioned about how he had done it. Subjects of the VI group all reported having pictured the objects before them, with varying degrees of vividness, and only 3 (of 15) said they had been bothered or confused by the locations of the objects in real space behind them. Their use of this imagery, however, apparently did *not* take the form of searching a mental picture to determine where

Table 2. Error data, Experiment I

Group	Errors		Number of subjects	
	Mean	Proportion	0 or 1 error	2 or more errors
Visual image	1.8	.041	10	5
Real space	3.2	.073	3	12
Control	1.3	.029	10	5

things were. The subjects agreed rather well in an account that went approximately as follows: as a question "On which side of the X is the Y?" was heard, X was pictured in its proper location, as soon as it was mentioned; whereas the critical word "Y" evoked a representation of the *location* of Y in imagined space (on the basis of which the subject could answer the question) *before* any picturelike representation of Y was formed. (It is entirely possible that X as well as Y was first located, then pictured, but that the subject had no reason to notice the order in the former case.) Twelve of the 15 subjects reported answering questions about orientation by "inspecting" the visual image of X; sometimes it was necessary to sharpen or clarify the image first. The other 3 subjects said that they depended on verbal memory for Orientation responses (e.g., 2 lists, one of objects facing or pointing left, the other of objects facing right). Subjects commonly told us that they verbalized both the locations and orientations of the objects in various ways *during the initial learning period,* but subsequently found these verbal devices too slow or clumsy to use on the recall task and quickly abandoned them in favor of imagery.

The reports of RS subjects were much more diverse. The most common strategy, reported by 6 of the 15 subjects, was one that we had never anticipated: it consisted of mentally translating the whole set of objects "over the head," so that they appeared in front of the subject as in a mirror, and answering questions in reference to this frontal construct in which left-right relations were the same (relative to the subject's new body orientation) as among the objects behind him. Five other subjects said that they imagined the objects in their true locations: 2 of these reported that the imagery was purely visual — like seeing with "eyes in the back of the head"; one said that it was rather a matter of "feeling" where the objects were, and 2 reported some mixture of visual and kinesthetic components. Two subjects, among the slowest in the group, reported predominantly verbal recall, and the remaining 2 were unable to tell us with any clarity how they had done the task.

Control subjects were interviewed less carefully; their reports closely resembled those of group VI.

EXPERIMENT II

The second experiment was essentially an improved version of the first. (1) In an effort to reduce "error" variability, a transfer design that allowed difference scores to be taken between a pretest and a posttest was adopted. (2) The people in group RS were instructed more emphatically

to keep imagining the objects in their real locations. The fact that 6 RS subjects in Experiment I did the task by means of a frontal mirror-image construct was interesting in its own right, but we were more interested in how people might imagine objects behind them without "cheating" in this way. (3) By this time we had become persuaded that the subjects' verbal accounts could tell us at least as much as their reaction times; therefore we questioned them more meticulously and exhaustively than before about the mental processes that the task involved, having learned from the earlier study what questions we wished to ask.

METHOD

Subjects

Forty-five new subjects were assigned to 3 equal groups as before.

Stimuli, apparatus, and procedure

The same array of stimulus objects was used, and every subject initially studied the display for 5 minutes, exactly as in Experiment I. Next, all subjects were treated like group C of Experiment I, responding to a full set of 44 questions about Location and Orientation while seated *facing* the objects, but with a curtain occluding the sight of them. This was the "pretest." The "posttest," consisting of the same 44 questions in a different order, was then given to group C under the same conditions (still facing the objects), whereas groups VI and RS were turned around with their backs to the array before taking it.

Instructions were recorded on tape and played back to the subjects. On first entering the experimental room, all subjects were instructed as follows:

If you look over at the shelf in front of you, you'll see a number of different objects: a *pipe,* with its stem on one side and its bowl on the other; a *duck,* with its head turned one way and its tail the other; a *pencil,* which has a point and an eraser at opposite ends; a *shoe,* with the heel on one side and the toe on the other side; likewise a *dog* with its head and tail, a *knife* with a handle and blade, and a *teapot* with a handle and a spout. I'd like you to sit here for a few minutes and study these objects. In the first place, try to remember where each of the objects is located, so that you could quickly point to it if you were blindfolded. Notice the position of each object relative to all the others. In the second place, try to remember how each of the objects is oriented: notice *which way* the knife is pointing, which way the duck is facing, and so on. It is *not* a good idea to memorize a *list* of the objects, because it takes time to refer to a mental list, and I'm going to be asking you some questions that will have to be answered very quickly. You need to remember the locations of many things in your everyday life; try to remember these objects in exactly the same natural way.

Go ahead, then, and look very carefully at *where the things are,* and *how they are oriented.*

Instructions for the pretest were also the same for all groups:

Now I'm going to ask you a series of questions about the objects. Most of the questions will take the form "On which side of the X is the Y?" — in which X

and Y are two different objects — and you can always answer the question by throwing the key-switch that's on the table before you either to the left or to the right. Reach out now and move the key left and right so that you get the feel of it... When you get through, leave the key in the middle position. Before each question I'll say "Ready," and when I do make sure that the key is in the *middle* position, and that you have your hand on it so that you're ready to throw it either left or right quickly and firmly.

Before we start, let me give you a couple of obvious examples. The pipe is on the far left, and the teapot is on the far right. Therefore, if I ask you "On which side of the pipe is the teapot?" you'll of course throw the key to the right. Or, *vice versa,* if I ask "On which side of the teapot is the pipe," you'll throw it to the left. And so on, likewise, for any other pair of objects.

Some other questions will test your memory for the way the objects are oriented. For instance, I might ask "On which side of the pencil is the pencil's point?" I won't tell you the answer, but if the pencil is pointing to the left you should throw the key to the left, and if it's pointing to the right you should throw the key to the right. You won't know in advance whether a question is going to be about a pair of objects, or about the way a particular object is oriented, because the questions are in no special order.

We want you to give *correct* answers, of course, but we also want you to give them as *quickly* as you possibly can without making a mistake. We're measuring your reaction time to each question, and the speed with which you answer is very important.

Now, do you have any questions about what you are supposed to do? Is everything clear?

After being turned around and reseated, subjects of group VI were told the following:

You are now sitting with your back to the objects, but I want you to imagine that you are still looking at them. Imagine that they are all laid out in front of you

Now, in your mental picture, the pipe should still be on the far left, and the teapot on the far right. Therefore, if I ask "On which side of the teapot is the pipe?" you should throw the key left, and so on. Likewise, if I ask "On which side of the pencil is the pencil's point?" you should throw the key left if the pencil was pointing left as you looked at it, and right if it was pointing right as you looked at it. Just give the answer that would be correct in terms of a picture of the objects.

If there's anything you need to have clarified before we go ahead, ask me about it now.

The corresponding instructions for group RS went as follows:

You are now sitting with your back to the objects; those that were on your left are now on your right, and *vice versa.* Imagine as clearly as you can where the objects are really located in relation to yourself as you are now sitting.

The teapot is on your far left behind you, and the pipe is on your far right behind you. Therefore, if I ask "On which side of the pipe is the teapot?" you should throw the key to the left, and so on. Likewise, if I ask "On which side of the pencil is the pencil's point?" you should throw the key left if the pencil is pointing to your left as it lies on the shelf behind you, and right if it is pointing to the right behind you. Just keep clearly in mind where the objects actually are in this room, and respond accordingly.

If there's anything you need to have clarified before we go ahead, ask me about it now.

Group C subjects were merely told informally at this point that they would be asked the questions a second time.

RESULTS

Pretest and posttest scores for individual subjects, in logarithmic form, are shown as scatter-plots in figures 1 and 2. Geometric means and change scores for the 3 groups are given in Table 3.

Note first that transfer from pretest to posttest was generally positive in direction, although knowledge of results was never supplied. Even group RS improved slightly in average RT, despite complete reversal of correct and incorrect responses between the first test and the second.

It does appear, however, that group RS improved somewhat less than did groups VI and C, at least on Orientation questions: $F(2, 42) = 3.85, .05 > p > .01$, for a comparison of the 3 distributions of change scores in logarithmic form, as in figures. The corresponding F for Loca-

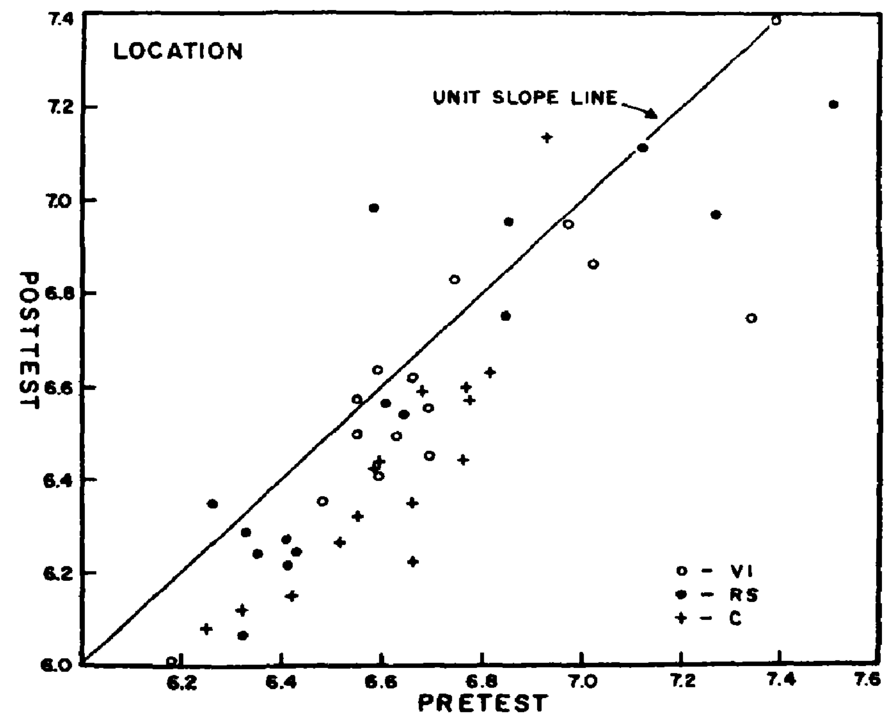


Figure 1. Posttest vs. pretest reaction times to questions about Location, Experiment II; each point shows mean natural-log reaction time of one subject

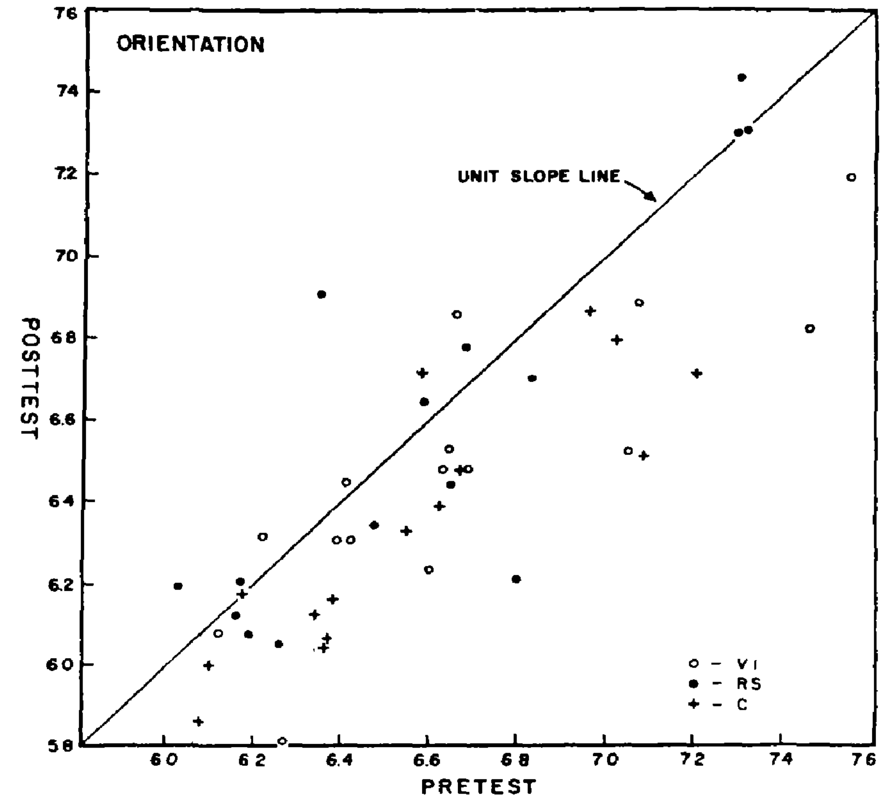


Figure 2. Posttest vs. pretest reaction times to questions about Orientation, Experiment II; each point shows mean natural-log reaction time of one subject

tion questions is 2.12, which is short of the .05 level; however, the configuration of change scores is similar for the 2 types of question, and comparison of change scores for Location between groups RS and C gives a $t(28) = 2.01$, which is at the .06 confidence level, considered in isolation. Our use of the pretest did contribute to the stability of the results, but residual within-group variability remains, unhappily, great (see figures) and the groups would still have to be quite large to provide much power for significance tests.

Error scores are shown both as means and as proportions in Table 4. Also tabulated are the number of subjects in each group who improved, stayed the same, or got worse from pretest to posttest, by a criterion of total errors. Further combining the "same" and "worse" categories gives a 2×2 table (with frequencies remarkably like those of Table 2 from Experiment I) for which $\chi^2(2) = 9.25, p < .01$. Further breakdown by Location and Orientation shows no interaction. Again, group differences

Table 3. Reaction times (geometric means, in msec) from Experiment II

	Type of question					
	Location			Orientation		
	Pre	Post	Change	Pre	Post	Change
Group						
Visual image	844	– 752	= 92	794	– 651	= 143
Real space	745	– 724	= 21	739	– 721	= 18
Control	748	– 615	= 133	711	– 569	= 142

are more reliably shown by errors (despite their small number) than by latencies, but ignoring the inconclusive RT results of Experiment I, time and error data show a consistent picture: groups VI and C about the same in performance; group RS worse, though not dramatically so.

The reaction time and error data are not inconsistent with the possibility of a response-reversal strategy on the part of group RS. When asked directly, however, people in this group unanimously denied having done the task in any such way; moreover, their accounts of what they did do were uniformly inconsistent with this possibility.

The "debriefing" interviews with group RS indicated that only 3 subjects employed the strategy, previously the most frequent, of translating the objects "over the head" to form a mirror-image array. (The revised instructions were intended to discourage this device, and may have done so.) Seven subjects — the modal subgroup — reported that they thought of the objects in their true locations, and in terms of purely visual imagery. An eighth subject joined this subgroup after attempting to use the frontal mirror-image device for the first 8 or 10 questions of the posttest. Another reported using a combination of visual and kinesthetic imagery of the objects behind him. Of the 3 remaining subjects, one reported doing the task by means that seemed purely kinesthetic (by "feel"), one — the slowest in the whole group — said that his coding of the objects was entirely verbal, and one — either less articulate or more prudent than the rest — was willing to say only that he "knew" where the things were.

Table 4. Error data, Experiment II

	Errors				Posttest vs. pretest errors: Number of subjects		
	Mean		Proportion				
Group	Pre	Post	Pre	Post	Better	Same	Worse
Visual image	1.53	.87	.035	.020	9	2	4
Real space	2.00	2.53	.045	.058	3	5	7
Control	2.07	1.07	.047	.024	11	1	3

The modal subjects insisted, under close questioning, that they were able to "see" the objects in place behind them. Several of them, comparing pretest and posttest, said it was harder to visualize objects behind than in front, or that the images were less clear, but no one suggested that they were different in kind. Asked if the experience was like that of turning the head or the body and looking at the objects behind, the subjects replied (with one or two ambiguous exceptions) that it was not; that they did not imagine turning the head from its real position, but rather looked at the objects as if through "eyes in the back of my head." This figure of speech was a recurrent one.

Since it is evident that group RS was composed of subgroups who were doing quite different things, one would like to know how these differed in RT and errors, but the small numbers make this almost impossible. The 7 subjects just discussed, who visualized the objects behind them, appear to have performed a little better than the group as a whole in terms of both time and errors, but not as well as groups VI and C. Two of these people had better error scores on posttest than on pretest, 3 had the same, and 2 worse; see Table 4. Comparison of this purified group RS with groups VI and C yields $\chi^2(2) = 3.96$, which is no longer significant, but with the reduced number the test is weak and, indeed, of dubious validity.

Interviews with the subjects of groups VI and C revealed very little variability in the way the task was done. The typical account detailed in the results section of Experiment I received repeated confirmation from the present subjects on all of its essential points.

DISCUSSION

At the outset we thought it likely: (1) that group VI might be confused by knowing where the objects really were, and that questions about Location (rather than Orientation) would be the more vulnerable to this interference; (2) that if group RS attempted to use visual imagery of the objects they had been viewing, left and right in the image would be reversed relative to the task demands, and that questions about Orientation were particularly likely to be answered by means of visual imagery and therefore to show decrement from this incompatibility. These expectations received little support from the results. Group VI was practically as good as the control; group RS showed some decrement, but apparently not for reasons that we had anticipated, and the effect was not appreciably different for Location and Orientation questions.

The subjects' reports on their mental processes turn out to be consider-

ably more informative than the objective results. On the "real space" task, people were able to employ visual imagery in 2 quite different ways, neither of which we foresaw. Some of them hit upon the trick of constructing a mirror-image array of the objects in front of them. (This rarely if ever involved imagining the objects actually reflected in a mirror; most people reported *moving the objects* to the front in a linear translation. Equivalently, they could have imagined moving themselves to an observation point behind the objects; this might have seemed more natural had the array not been at the extreme end of the room.) Although the subjects were in a sense "cheating" — i.e., not really imagining the objects behind them, as they had been instructed to do — this technique at least demonstrates the remarkable facility with which visual images can be manipulated and transformed.

However, a considerable number of people given the RS task, particularly in Experiment II, *did* imagine the objects behind them, and furthermore maintained that the imagery was *visual*, though unaccompanied by imagery of turning-and-looking. This contradicts the view that a visual image must correspond to some possible visual input. (But otherwise what makes it visual?) We are being told, in effect, that the *mind's* eye has a cycloramic, 360-degree field. Most of the subjects said (consistently with time and error results) that imagery localized in the rear was generally poorer than that in front, but that there was no discontinuity between one and the other.

One of the authors (Attneave, 1974) recently suggested a representational model in which a hierarchical descriptor structure, serving to identify objects, is connected by a system of "place-markers" to a tridimensional analogue medium representing physical space, in which the objects are thereby located. A place-marker from a high level of the hierarchy might indicate gross location of a whole object, but a configuration of markers from the lowest levels — descriptors of, say, points and lines — would essentially *copy* an object mapped into the space from sensory input, or could, alternatively, constitute the *image* of an object without sensory support. Perception of a stable world would require close tracking of input by place-markers in the case of head and eye movements, and this conceptualization is quite compatible with the well-known theory of von Holst (1954). The fact that more detail can be "seen" in peripheral vision than is supplied by the retina (Hochberg, 1968) may be understood in terms of place-marker configurations that are created in central vision and transposed to the periphery when the eyes move. But suppose that a head or eye movement takes an object out of the field altogether: what then becomes of its place-markers? They must disappear, having no place to go, if the representational space has the same boundaries as the visual

field. If this space is not so bounded, however, but extends to represent the world behind the head, then configurations of place-markers (i.e., images) might be transposed from the region of their genesis into a sector that receives no sensory input at all. The *visual* quality of the latter (and our subjects' experience of seeming to see through eyes in the backs of their heads) would be accounted for by the continuity of the representational medium that allows transposition of patterns between the region that receives visual input and that which does not (cf. Attneave and Olson, 1971). Although the essentials of this model were developed earlier, it was not until the present studies were done that we appreciated the possibility of a cycloramic visual field. Without being uncritical of the purely phenomenal evidence that supports it, we now believe this possibility to be a serious and plausible one.

How the subjects *used* visual imagery (whether it was localized in front or behind) deserves a little further consideration. In no case did anyone claim to have held a detailed picture of all the objects at once while doing the task. Rather, the name of an object *evoked* its image, in a determinate spatial region, often with some visual imagery of proximate objects. If the questions concerned Orientation, the image then seemed to supply the answer directly. If the question concerned Location, however, the second object was located in space (subjects typically said) and the question was answered accordingly, before the object was clearly pictured. This is logically consistent with the report that images were evoked or constructed, rather than continuously present: one must decide *where* to draw a picture before drawing it. Are we to interpret this account, however, to mean that visual images did *not*, after all, contribute to the answering of Location questions?

The issue here is perhaps one of definition rather than of substance. Information about the internal structure of the objects was simply irrelevant to questions of Location. On the other hand, even two labeled points (standing for the represented objects), located in a representational space, constitute a rudimentary picture. In the terms of the model sketched above, we suggest that the difference is merely one of the level of a hierarchical descriptor-structure to which the place-markers belong.

Notes

This research was supported by a grant from the National Institute of Mental Health, MH 20449-03, for "Studies on Spatial Representation." Requests for offprints should be sent to Fred Attneave, Psychology Department, University of Oregon, Eugene, OR 97403. Received for publication October 19, 1976; revision, January 11, 1977.

References

Attneave, F. 1972. The representation of physical space. In *Coding processes in human memory,* ed. A. W. Melton and E. J. Martin. Washington, D.C.: Winston.

Attneave, F. 1974. Apparent movement and the what-where connection. *Psychologia* 17:108–120.

Attneave, F., and Olson, R. K. 1971. Pitch as a medium: A new approach to psychophysical scaling. *American Journal of Psychology* 84:147–166.

Hochberg, J. 1968. In the mind's eye. In *Contemporary theory and research in visual perception,* ed. R. N. Haber. New York: Holt, Rinehart, and Winston.

Piaget, J., and Inhelder, B. 1956. *The child's conception of space.* London: Routledge and Kegan Paul.

Tolman, E. C. 1948. Cognitive maps in rats and men. *Psychological Review* 55: 189–208.

von Holst, E. 1954. Relations between the central nervous system and the peripheral organs. *Animal Behavior* 2:89–94.

Consultants

The editors of the *American Journal of Psychology* take this opportunity to thank those who have recently assisted in the evaluation of manuscripts.

Paul Bakan
Michael Birnbaum
Richard G. Bryan
George Collier
Clyde H. Coombs
Alice Hendrickson Eagly
Ward Edwards
Charles Flaherty
Donald J. Foss
Ronald G. Gandelman
William C. Gordon
William W. Grings
Ervin Hafter
Steven J. Haggbloom
Leonard Hamilton
Robert Henderson
Edward Hirsch
Lawrence Jones
Tracy S. Kendler
Patrick Laughlin
Robert Lehr
Alan Leshner
Charles Lewis
Stephen Lisman
James L. McClelland
William J. McGuire
D. J. K. Mewhort
Peter J. Mikulka
Ulric Neisser
Raymond Romanczyk
John J. Shaughnessy
Edward J. Shoben
Dan I. Slobin
David Sparks
Pamela Steinert
Endel Tulving
Craig Twentyman
Charles Watson

AMERICAN JOURNAL OF PSYCHOLOGY
December 1977, Vol. 90, No. 4, pp. 565–575

Task difficulty as a mediator of subject strategy in memory scanning

John C. Schmitt and C. James Scheirer
State University of New York at Binghamton

A number of previous studies have shown performance to be mutable in the Sternberg memory-scanning task. Our results suggest that the difficulty of the task, and particularly of the list materials employed, may underlie a number of findings at variance with a strict scanning interpretation. Experiment I found that probes of letter-sets that form words are responded to faster than those of nonwords, irrespective of the number of list items. Experiment II replicated these findings and extended the investigation to the influence of practice in this situation. We conclude that criterion changes in the response-decision process provide a straightforward model for reaction-time effects when experimental conditions are perceived as differentially difficult.

The experimental paradigm and theory developed by Sternberg (1969) concerning the dynamics of recognition performance have been extensively investigated in the past decade. A striking aspect of results in this area has been the extent to which basic processes of memory scanning can be modified to fit the special demands of the task at hand. For example, Sternberg proposed that scanning is exhaustive, rather than self-terminating. He argued that the scanning process is so fast that it is usually more efficient to scan all items than to stop and check after each. However, when scanning was for location, rather than for presence of an item, subjects were able to modify the scanning process to reflect the greater relative efficiency of a self-terminating strategy (Sternberg, 1969, Exp. 7).

A number of other investigators have found that performance in the Sternberg task is modifiable by subject strategies. Mohs, Westcourt, and Atkinson (1975), for example, found that, after extended (3–4 days) practice, search strategies for items in a categorized list changed from hierarchical to associative. Atkinson and Juola (1974) present convincing evidence that a memory-scanning strategy is used only when the familiarity of a probe is not high or low enough to allow immediate re-

sponding. Burrows and Okada (1976) found that physical and category representations of words may be accessed and scanned independently, in order to facilitate performance. Schmitt and Scheirer (1977) found that lists may be segregated into separate digit and consonant subsets and scanned independently when conditions suggest this to be most efficient strategy.

The purpose of the present report is to investigate "perceived task difficulty" as a mediator of subject strategies. In particular, some results by DeRosa and Morin (1970) suggest to us that the difficulty of the memory load (list being scanned) has an influence on retrieval strategy. They investigated the scanning of digit sets where the digits were either consecutive or nonconsecutive. Although the experiment was designed to investigate serial position effects in the Sternberg task and only a single list length (four digits) was used, probes of consecutive sets were responded to an average 24 msec faster than those of nonconsecutive ones. Among the various consecutive sets, the list "1-2-3-4" was responded to most rapidly. In an earlier experiment, Morin, DeRosa, and Stultz (1967) found that probes of nonconsecutive sets that could be reordered into a consecutive set were responded to 33 msec faster than those of "open" sets.

It seems clear to us that difficulty is a parameter of all cognitive tasks, not just the Sternberg one. A recent theoretical paper by Baddeley and Hitch (1974) suggests many experimental paradigms are interactive in the sense of reducing or slowing down performance when performed simultaneously (see their review). Therefore, a search for the common mechanisms across tasks will lead to understanding of the system rather than understanding of only the paradigms. If perceived task difficulty is a general mediator of working-memory strategies, the results of DeRosa and Morin, and of Morin, DeRosa, and Stultz, should be replicable in form with other variables presumed to reflect task difficulty. We manipulated difficulty through the use of lists of letters that formed either words or various types of nonword sets in a memory-scanning task. Experiment I investigated the effects of such sets in a between-groups design; Experiment II replicated the results of Experiment I and extended the design to look at changes in strategy due to practice effects and the switching of material type within a single session.

EXPERIMENT I

METHOD

Stimulus materials

Presentation sets of 3, 4, 5, and 6 capital letters were xeroxed on plastic film and projected on a screen 2 meters in front of the subject. Letter-set conditions

were chosen to form an apparent continuum of difficulty. Word sets were randomly drawn sets of length 3, 4, 5, or 6 letters that formed words; half of the words had a frequency of 25–50 per million and half had a frequency of 2 (Kucera and Francis, 1967). The words chosen were all English words; no numbers, abbreviations, contractions, or proper names were allowed. The words contained no repeated letters and could be rearranged to form the pronounceable and nonpronounceable control sets. This latter restriction resulted in the exclusion of approximately 25% of the words selected; this exclusion led to no obvious selection bias in the words retained.

Pronounceable letter sets were constructed by rearranging the letters in word sets such that a pronounceable nonword resulted (e.g., TRIP-PRIT). Nonpronounceable letter sets were constructed by rearranging the letters in the word sets such that a nonpronounceable nonword was formed (example: TRIP-RTPI). Random-letter sets were constructed by randomly selecting sets of 3, 4, 5, and 6 letters from the letters of the alphabet without replacement.

Probe letters were randomly drawn on the 50% of trials that were negative, with the restriction that the selected letter not be a member of the letter set; on positive trials probes were balanced for serial position in the set. There were 2 completely independent replications of each condition with different lists.

Procedure

Each subject received 96 experimental trials, preceded by 8 nontabulated practice trials. These 96 trials were randomly distributed, in blocks of 32 trials, among 16 within-subject conditions. These conditions represented the factorial combination of response type (positive or negative), frequency (high or low) and set size (3, 4, 5, and 6). Frequency in the pronounceable and nonpronounceable conditions was determined by the frequency of the root words of the words condition. In the random-letter conditions, frequency was assigned randomly to sets in order to complete the factorial design. Subjects were instructed to study the letter sets for as long as they wished, and to indicate when they were done studying by saying "ready." Approximately 3 seconds later a single probe letter was presented and an appropriate (yes-no) button press required. After making the responses, subjects recalled the list letters in the order in which they were presented in order to insure attention to the task and to discourage unscrambling of the letters. Three seconds later a new list was presented, followed by a new probe. Responses were made by depressing response buttons labeled "yes" and "no." The positions of these buttons were counterbalanced such that for one-half of subjects the preferred hand signaled "yes," while for the remainder of subjects the nonpreferred hand signaled "yes." Positive and negative responses were made by pressing these buttons with index fingers.

Subjects

Sixty-four introductory-psychology students at the State University of New York at Binghamton served as subjects in order to fulfill a course requirement. Assignment to the various letter-set conditions and counterbalancings was determined by a random schedule. All subjects were naïve both to the experimental paradigm and the nature of the sets employed.

RESULTS

Harmonic-mean latencies of responses were computed on the 6 trials in each condition for each subject. The rationale for the use of harmonic-mean rather than arithmetic-mean latencies has been discussed by Scheirer and Hanley (1974). Errors of response accounted for 1.7% of the trials overall and error trials were excluded from the latency analysis. The proportions of these errors, tabulated by list condition and set size are presented in Table 1, and do not indicate any speed-accuracy trade-off in these data. Latencies greater than 2.5 seconds accounted for .29% of the data and were also discarded from the analysis.

The linear trend relating the number of letters in the sets and reaction time was highly significant, $F(1, 60) = 230.42, p < .01$. This linear component accounted for 98.1% of the variability due to set size; there was no interaction of linear trend and list conditions, $F(3, 60) < 1$.

The type of material in the set had an overall effect on reaction time, $F(3, 60) = 2.78, p < .05$, with the comparison of word sets with the 3 other list types accounting for 91% of this effect, $F(1, 60) = 7.55, p < .01$. The pronounceable, nonpronounceable, and random letters conditions did not differ, $F(2, 60) < 1$. These effects are illustrated in Figure 1.

Response type ("yes" or "no" response) had no overall effect, $F(1, 60) < 1$, nor did it interact with letter-set conditions, $F(3, 60) = 1.18, p > .05$. This latter finding is consistent with Sternberg's (1969) hypothesis of exhaustive search of the sets in this situation, as the slopes for positives and negatives would differ were the search to terminate following location of the item.

Frequency of the words from which the word, pronounceable, and nonpronounceable sets were constructed had no effect on reaction time, $F(1, 60) = 2.37, p > .05$ and there were no significant 2-way interactions involving frequency. Due to this lack of effects, the frequency variable was not included in the design of Experiment II.

Table 1. Mean percentages of incorrect responses by set size and letter-set condition in Experiment I

	Set size			
Letter-set condition	3	4	5	6
Words	1.56	1.56	.52	2 08
Pronounceables	1.04	52	.78	1.04
Nonpronounceables	1.82	2.60	2.86	1.30
Random letters	1.56	3 90	1.82	2.60

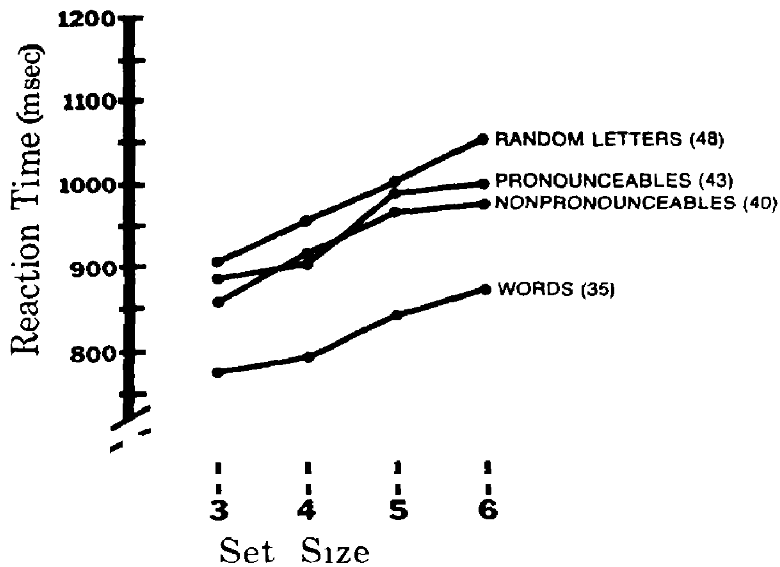


Figure 1. Mean reaction time to probes of letter sets in Experiment I. Slopes of the best-fitting straight lines are in parentheses

DISCUSSION

These findings are consistent with the serial exhaustive scanning model of performance outlined by Sternberg (1969). This model states that overall response latency reflects the time taken by a series of additive processing stages; these processing stages include encoding of the probe, serial comparison of the probe with each of the items in memory, response decision, and organization of the response itself. The linear slope of reaction time across set size is taken as evidence for the serial nature of the comparison process (scan). Identical slopes for positive and negative probes indicate that the scan is exhaustive, that is, that all items in the list are scanned.

The potential sensitivity of the Sternberg memory-scanning paradigm in detecting shifts in retrieval strategy lies in its multistage assumptions. If retrieval is seen as the result of independent processing stages, then any stage might be sensitive to cognitive control. While there has been considerable disagreement as to the validity of Sternberg's specific model (cf. Theios, Smith, Haviland, Traupmann, and Moy, 1973, and Atkinson and Juola, 1974) we regard it as a useful framework within which to discuss memory processing.

In the Sternberg scanning model, either slopes or overall levels of re-

action time may be affected by experimental manipulation. Changes in slope reflect changes in processes occurring in the scanning stage, and changes in the overall level reflect changes in one or more of the other stages. The results of Experiment I showed that probes of words were responded to faster than probes of nonwords, while the slopes across set size did not differ. This indicates noninvolvement of the scanning stage under these conditions. It is unlikely that the probe-encoding stage is affected by set type, since probe stimuli did not differ between letter-set conditions and should therefore be "refined" with equal speed. It is also unlikely that the response-execution mechanism was affected since response probabilities did not differ among conditions. If we may assume that the stage analysis conducted by Sternberg (1969) is complete with respect to the stages of this task, only the binary decision stage remains to account for the overall level differences.

In the binary decision stage it is assumed that the subject uses some product of the scanning stage to make a "yes" or "no" decision as to the presence of the probe in the memorized letter-set. As a tentative explanation, it is possible that low confidence due to the difficulty of the set might raise the criterion for detection of the "signal." This would result in faster reaction times for easier sets, independent of slope differences. A criterion-shift notion would indicate a general tendency to maintain execution of the appropriate cognitive processes but increase caution at the "checkpoints" (decision nodes) of information flow. As such, it would be consistent with speed-accuracy trade-off phenomena, where increased speed, potentially indicating lesser criteria at decision stages, has been associated with decreased accuracy of performance (cf. Pachella, 1974).

Experiment II was concerned with the further investigation of possible difficulty-related criterion shifts. Practice at the task was examined using 3 of the letter-set conditions of Experiment I. In addition, the design of Experiment II afforded a replication of the word-nonword effects found in Experiment I.

EXPERIMENT II

METHOD

Subjects

Seventy-two introductory-psychology students at the State University of New York at Binghamton served as subjects in order to fulfill a course requirement. Assignment to the various list conditions and counterbalancings was determined by a random schedule. All subjects were naïve both to the experimental paradigm and the nature of the sets employed.

Procedure

All apparatus, time parameters, and counterbalancing procedures were the same as those of Experiment I. There were 9 between-subject conditions representing the factorial combination of either words, nonpronounceables, or random letters in the first half of the session; and words, nonpronounceables, or random letters in the second half of the session. For the 6 conditions where a switch of material type was involved, the subject was not informed or given any signal that a change of material was to occur. All stimulus sets and probe letters were randomly selected from those used in Experiment I and there were a total of 128 experiment trials for each subject. No practice trials were given for any letter-sets conditions.

RESULTS AND DISCUSSION

Harmonic-mean reaction times were computed on the 8 trials in each condition for each subject. Errors of response accounted for 1.79% of the data and were excluded from the analysis. Error rates by letter-set condition, set size, and block of trials (first half vs. second half) are reported in Table 2. Latencies greater than 2.5 seconds accounted for 0.65% of the data and were also discarded.

It should be noted that the presentation of words, nonpronounceables, and random letters to unpracticed subjects provides a replication of Experiment I except that the pronounceable letter-set condition was eliminated to reduce design complexity. The results of the replication may be seen in the left panel of Figure 2. A simple-effects analysis of the first-half data revealed that reaction times to probes of words were signifi-

Table 2. Mean percentages of incorrect responses by set size and letter-set condition in the first and second halves of Experiment II

	First Half			
	Set size			
Letter-set condition	3	4	5	6
Words	1.56	1.30	2.86	1.04
Nonpronounceables	1.56	.78	1.30	2.08
Random letters	1.30	1.56	2.08	4.16
	Second Half			
	Set size			
Letter-set condition	3	4	5	6
Words	1.30	.78	1.30	.52
Nonpronounceables	.26	.52	1.30	2.86
Random letters	1.56	2.08	5 20	3.64

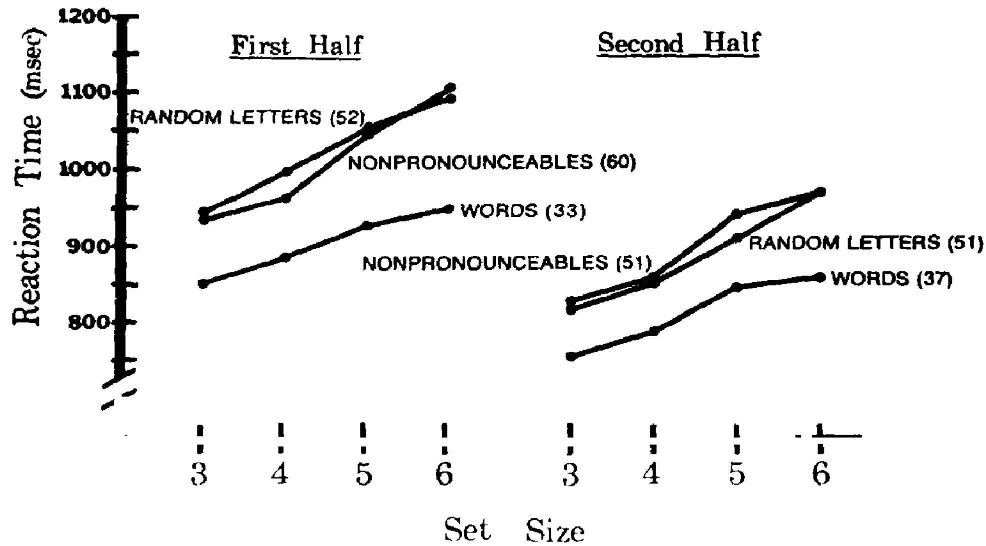


Figure 2. Mean reaction time to probes of letter sets in the first (left section) and the second (right section) halves of Experiment II; slopes of the best-fitting straight lines are in parentheses

cantly faster than to those of nonpronounceables and random letters pooled, $F(1, 69) = 5.33, p < .01$. Nonpronounceables and random letters did not differ, $F(1, 69) < 1$. This result replicates the word vs. nonword effect found in Experiment I. In addition, greater slopes across set size for the more difficult (nonword) conditions appeared, $F(1, 69) = 9.40$, $p < .01$ with slopes of 33 vs. 56 msec per item. This effect was nonsignificant in Experiment I, although the slopes were in the correct direction (35 vs. 44 msec per item), and was marginally significant in the second-half analysis of Experiment II, $F(1, 63) = 3.00, .10 < p < .05$ with slopes of 37 and 51 msec per item. These results taken together indicate that difficulty may involve the scanning stage of memory processing to a minor extent.

The most striking result noted in Experiment II was the effect of practice. Reaction times in the second half of the experimental session were an average of 112 msec faster than those in the first half, $F(1, 63) = 87.02, p < .01$. This practice effect did not interact with the slopes of reaction times across set size, $F(1, 63) < 1$, and therefore appears to have had a strictly additive effect on these functions. In addition to the simple effects of practice, it was found that the magnitude of practice effects depended on the particular order of the letter-set conditions. When subjects were switched from more-difficult to less-difficult material, very large

practice effects appeared, while changes from easy to hard materials produced minimal practice effects. As may be seen in Table 3, the magnitude of the practice effects varied positively with the difficulty of the first-half material type, and inversely with the difficulty of the second-half material type. The pattern of effects did not differ between the first and second half of the experiment, $F(4, 252) < 1$.

The usual results of the Sternberg paradigm were also noted. The linear component of trend across set size, averaged across the entire experiment, was significant, $F(1, 63) = 207.80, p < .01$, and accounted for 98.7% of the variability due to set size. The slopes of reaction times across set size again did not differ with response type (positive or negative), $F(1, 63) < 1$.

CONCLUSIONS

The earlier research by DeRosa and Morin and by Morin, DeRosa, and Stultz indicated that reaction-time effects may be correlated with the apparent difficulty of lists in the Sternberg paradigm. Practice effects in this task are also well known (e.g., Krueger, 1975; Simpson, 1972). Additionally, some data collected by Hoyer, Connor, and Scheirer (in press) suggest that the addition of a concurrent memory load also increases response latency relative to a no-load condition. Similar results have been reported by Darley (Note 1) and Kaminsky and DeRosa (1972). The strong communality of all of these findings with the present wordness and practice effects suggest that they, and perhaps others, are entirely explainable as due to difficulty-related criterion shifts.

Sternberg's model of scanning provides no basis on which reaction times for letters in "difficult" sets should differ from those in less-difficult sets of equal size. This was not our result; in fact, it appears that a constant delay for all set sizes was added by the difficulty manipulation. There is

Table 3. Magnitudes of practice effects (mean first-half response latency minus mean second-half response latency), in milliseconds, by letter-set condition in Experiment II

	Second-half letter-set condition		
	Words	Nonpronounceables	Random letters
First-half letter-set condition			
Words	168	61	−8
Nonpronounceables	149	126	92
Random letters	176	140	116

a strong suggestion of active subject strategies in these data. Specifically, it is possible that delayed processing occurs at the point of response decision as a result of the perceived difficulty of the lists. Increasing difficulty of the list might result in a strategy of longer sampling of whatever signal from the scanning stage indicates a match or nonmatch of the probe with the memorized list. This would be sufficient to cause the overall reaction time differences noted for word/nonword comparisons in both experiments and for practice in Experiment II.

As indicated earlier, criterion-shift notions of performance extend outside the Sternberg paradigm as well, particularly in the area of speed-accuracy trade-off in reaction time. Both areas illustrate the inherent flexibility and adaptability of subject strategies. Indeed, it may be argued that alternative models of scanning (e.g., the suggestion of Theios, Smith, Haviland, Traupmann and Moy (1973) that with intensive practice on fixed lists subjects scan in a self-terminating mode) reflect this mutability of processing rather than the nonvalidity of Sternberg's model.

Notes

Experiment I of this paper was part of the master's thesis of the first author. The authors wish to thank Eric Goldstein and Emile Shulman for their assistance in the execution of Experiment II. Requests for reprints should be addressed to C. James Scheirer, Department of Psychology, State University of New York at Binghamton, Binghamton, NY 13901. Received for publication September 13. 1976.

Reference Notes

1. Darley, C. F. Effects of memory load and organization on short-term recognition memory. Doctoral dissertation, Stanford University, 1973.

References

Atkinson, R. C., and Juola, J. F. 1974. Search and decision processes in recognition memory. In *Contemporary developments in mathematical psychology*, ed. D. H. Kroutz, R. C. Atkinson, and P. Suppes. San Francisco: Freeman.

Baddeley, A. D., and Hitch, G. 1974. Working memory. In *The psychology of learning and motivation: Advances in research and theory, Vol. 7*, ed. G. Bower. New York: Academic Press.

Burrows, D., and Okada, R. 1976. Parallel scanning of physical and category information. *Memory & Cognition* 4:31–35.

DeRosa, D. V., and Morin, R. E. 1970. Recognition reaction time for digits in consecutive and nonconsecutive memorized sets. *Journal of Experimental Psychology* 83:472–479.

Hoyer, R. G., Connor, J. M., and Scheirer, C. J. (In press.) Effects of memory load on short-term recognition latency. *American Journal of Psychology.*

Kaminsky, C. A., and DeRosa, D. V. 1972. Influence of retrieval cues and set organization on short-term recognition memory. *Journal of Experimental Psychology* 96:449–454.

Krueger, L. E. 1975. The effect of an extraneous added memory set on item recognition: A test of parallel-dependent vs. serial-comparison models. *Memory and Cognition* 3:485–495.

Kucera, H., and Francis, W. N. 1967. *Computational analysis of present-day American English.* Providence: Brown University Press.

Mohs, R. C., Westcourt, K. T., and Atkinson, R. C. 1975. Search processes for associative structures in long-term memory. *Journal of Experimental Psychology: General* 104:103–121.

Morin, R. E., DeRosa, D. V., and Stultz, V. 1967. Recognition memory and reaction time. *Acta Psychologica* 27:298–305.

Pachella, R. G. 1974. The interpretation of reaction time in information-processing research. In *Human information processing: Tutorials in performance and cognition,* ed. B. Kantowitz. Hillsdale, N.J.: Erlbaum.

Scheirer, C. J., and Hanley, M. J. 1974. Scanning for similar and different material in short- and long-term memory. *Journal of Experimental Psychology* 102:343–346.

Schmitt, J. C., and Scheirer, C. J. 1977. Segregation strategies in memory scanning. *American Journal of Psychology* 90:39–44.

Simpson, P. J. 1972. High-speed memory scanning: Stability and generality. *Journal of Experimental Psychology* 96:239–246.

Sternberg, S. 1969. Memory-scanning: Mental processes revealed by reaction-time experiments. *American Scientist* 57:421–457.

Theios, J., Smith, P. G., Haviland, S. E., Traupmann, J., and Moy, M. C. 1973. Memory scanning as a serial, self-terminating process. *Journal of Experimental Psychology* 97:323–336.

continued from page 736

Murch, Gerald M., ed. *Studies in Perception.* Indianapolis: Bobbs-Merrill, 1976. Pp. 364. Paperback.

Murphy, Lois Barclay, and Moriarity, Alice E. *Vulnerability, Coping, and Growth: From Infancy to Adolescence.* New Haven, Conn.: Yale University Press, 1976. Pp. 460. $20.00.

Myers, Patricia I., and Hammill, Donald L. *Methods for Learning Disorders.* 2d ed. New York: Wiley, 1976. Pp. 447.

Newson, John, and Newson, Elizabeth. *Seven Years Old in the Home Environment.* New York: Halsted, 1976. Pp. 436.

Novak, David. *Suicide and Morality.* New York: Scholars Studies Press, 1976. Pp. 136. $7.50.

Nye, F. Ivan. *Role Structure and Analysis of the Family.* Beverly Hills, Calif.: Sage Publications, 1976. Pp. 224. Paperback. $7.00.

Obudho, Constance E. *Black-White Racial Attitudes: An Annotated Bibliography.* Westport, Conn.: Greenwood Press, 1976. Pp. 180. $11.50.

Oden, Thomas C. *The Transactional Awareness Game.* New York: Harper & Row, 1976. $15.00.

Orford, Jim. *The Social Psychology of Mental Disorder.* New York: Penguin Books, 1976. Pp. 266. Paperback.

Ornstein, Robert E. *The Mind Field.* New York: Viking Press. 1976. Pp. 141. $7.95.

O'Shea, John Patrick. *Scientific Principles and Methods of Strength Fitness.* 2d ed. Reading, Mass.: Addison-Wesley, 1976. Pp. 192. Paperback.

Osman, Jack D. *Thin from Within.* New York: Hart, 1976. Pp. 400. $9.95.

Oyle, Irving. *Time, Space and the Mind.* Millbrae, Calif.: Celestial Arts, 1976. Pp. 145. Paperback. $4.95.

Pace, J. Blair. *Pain, A Personal Experience.* Chicago: Nelson-Hall, 1976. Pp. 163. $7.95.

Paranjpe, A. C. *In Search of Identity.* New York: Halsted, 1976. Pp. 263. $12.50.

Park, Clara Claiborne, and Shapiro, Leon N. *You Are Not Alone.* Boston: Little, Brown, 1976. Pp. 496. $15.00.

Pascal, Harold. *The Marijuana Maze.* Canfield, Ohio: Alba Books, 1976. Pp. 113. Paperback. $1.75.

Patterson, Gerald R. *Living with Children.* Champaign, Ill.: Research Press, 1976. Pp. 132. Paperback. $3.95.

Perkowski, Jan C. *Vampires of the Slavs.* Cambridge, Mass.: Slavica Publishers, 1976. Pp. 294. Paperback. $6.95.

Perlman, Helen Harris. *Persona: Social Role and Personality.* Chicago: University of Chicago Press, 1976. Pp. 245. Paperback. $4.50.

Peto, Frank. *Cosmic Psychology.* Houston: Yates, 1976. Pp. 130. Paperback.

Piaget, Jean. *The Grasp of Consciousness.* Cambridge, Mass.: Harvard University Press, 1976. Pp. 360.

Pieraut-Le Bonniec, Gilberte, and Van Meter, Karl. *Etude Genetique de la Construction d'une Propriete Relationnelle.* Paris: Centre National de la Recherche Scientifique, 1976. Pp. 98. Paperback.

continued on page 608

AMERICAN JOURNAL OF PSYCHOLOGY
December 1977, Vol. 90, No. 4, pp. 577–586

Contextual effects in duration experience

Douglas J. Bobko, H. R. Schiffman, Robert J. Castino, and Wayne Chiappetta
Rutgers — The State University

The role of contextual factors on duration estimates was investigated, employing 6 time intervals ranging from 15 to 35 sec (demarcated by the onset and termination of a display panel of lights). When compared with earlier research, the results suggest that duration estimates are affected by the context of the stimulus intervals with regard to other stimuli in the series. Specifically, those stimuli that were overestimated when they were the shortest members of the series were underestimated when they were the longest intervals of the stimulus series. In addition, a lengthening effect was observed: duration estimates increased over blocks of trials for all stimulus intervals.

The effects of an experimentally imposed context on the judged characteristics of a given stimulus have received considerable attention (e.g., Helson, 1964; Parducci, 1974). Typically, judgments of a stimulus have been shown to be dependent not only on the absolute qualities of the stimulus, but on the range, frequency, magnitude, *inter alia,* of other stimuli in the series as well. For example, the introduction into the stimulus series of an anchor stimulus may result in either the assimilation of judgments toward the anchor (Postman and Miller, 1945; Goldstone, Lhamon, and Boardman, 1957) or a contrast effect (Behar and Bevan, 1961), depending in part on the distance of the anchor from the average value of the stimulus series.

In temporal experience, one of the more common and stable contextual effects is defined as Vierordt's Law: Within a given series of stimulus intervals, the longer intervals tend to be underestimated and the shorter intervals tend to be overestimated (see Woodrow, 1951, p. 1225; Underwood, 1966, p. 57). In fact, these systematic "errors" in duration estimates exemplify Hollingworth's (1910) more general principle of the "central tendency of judgment"; that is to say, judgments of the characteristics of stimuli tend to assimilate or regress toward the value of the median or

mid-range stimulus in the stimulus set such that small differences between a given stimulus and the mid-range of the stimulus series are minimized (see Fraisse, 1963, pp. 116–128, for a review).

In a previous study investigating the effects of stimulus complexity on duration experience (Schiffman and Bobko, 1974), the findings confirmed Vierordt's Law for a series of 6 stimulus intervals of 3 to 23 seconds in 4-sec increments; subjects tended to overestimate the "shorter" stimulus intervals (3 and 7 seconds) and underestimate the "longer" stimulus intervals (19 and 23 seconds). However, since the majority of earlier studies reporting regression effects employed no stimulus interval longer than several seconds, the applicability of Vierordt's Law to longer intervals remains unclear. Accordingly, the present study attempted to assess the generality of Vierordt's Law for longer intervals by changing the stimulus series to 15 to 35 seconds in increments of 4 seconds. Moreover this study replicated and elaborated the procedures of Schiffman and Bobko; specifically, the stimulus intervals 15, 19, and 23 seconds were the shortest intervals of the present stimulus series but the longest intervals of the previous study. According to Vierordt's Law, duration estimates should be influenced by contextual factors and not simply the absolute magnitude of the stimuli themselves. That is, the "long" stimulus intervals that were underestimated in the previous study should be overestimated when, relative to the other stimuli in the present series, the intervals are "short."

The present experiment also allowed an assessment of the effects of practice on duration estimates. The central tendency principle predicts that systematic errors of estimation become more apparent as familiarity or practice with the stimulus series increases, i.e., the under- and overestimation of long and short intervals, respectively, should become more pronounced over blocks of trials. However, Eson and Kafka (1952) and Falk and Bindra (1954) did not report such an effect for the method of production and found, rather, that duration productions increased for all intervals of a series over the course of the experiment; Treisman (1963) also reported a general tendency for estimates to increase over trials for both the methods of production and reproduction, and he termed this trend the "lengthening effect." An examination of practice effects in the present study, then, was intended to provide some clarification of this issue; given the conflict between theoretical predictions and scant experimental data, no specific hypotheses were formulated concerning the effects of practice on duration estimates. Of additional interest was the replication of the main effects of stimulus complexity on duration estimates (Schiffman and Bobko, 1974), i.e., the greater the level of

complexity (as defined herein) occurring within a given stimulus interval the longer the estimate of the interval.

METHOD

Subjects

The subjects were 63 volunteers (36 female, 27 male) from an introductory course in psychology. All subjects, although aware that the task was of time estimation, were naïve with regard to the experimental hypotheses.

Stimuli and apparatus

The experimental stimuli consisted of 3 light displays, each display consisting of 2 rows of 4 lights (see Schiffman and Bobko, 1974, and Schiffman and Hamilton, 1974, for a more complete description). The duration of the lighting of each display was controlled by an interval timer in circuit with each panel display. A *low-complexity* display panel was wired so that all lights had the same onset and remained continually lit until a specified interval had ended. An *intermediate-complexity* display panel was wired to a rotating cam of relays so that all lights flashed on and off simultaneously and at a constant rate (2 flashes per sec) for the duration of the interval. This display produced a regular and predictable lighting pattern. A *high-complexity* display panel was constructed that enabled the 8 lights to flash on and off in a random sequence.

The responses — time estimations — were made by the method of reproduction (see Hornstein and Rotter, 1969). A microswitch, connected to a digital stop clock, was mounted on a small table beside the subject. Pressing of the switch started the clock and release of the switch stopped the clock. The subjects were instructed in using the switch to produce their responses.

All recording and control equipment were in a room adjacent to the testing room. Room lighting was about 79 cd/m^2. Beneath each panel were two instruction lights. A 2-sec flash of an amber light preceded the onset of the display and served as a "ready" signal; it indicated that the lighting display was about to be activated. A blue light was a "respond" signal. It signaled the subject to respond by pressing the response switch, and it remained lit until 2 sec after the response had ended.

Design and procedure

There were 3 independent groups of 21 subjects each. The subjects were assigned randomly to one of three levels of complexity: *low, intermediate,* and *high*. The 2 other variables were stimulus duration and block of trials. Within each level of stimulus complexity each subject was presented with 3 blocks of intervals; each block contained the intervals 15, 19, 23, 27, 31, and 35 seconds so that each subject made 3 estimates per stimulus interval. These intervals were selected so that the 3 shortest intervals were equal to the 3 longest intervals of the previous experiment (Schiffman and Bobko, 1974). The intervals in each block were presented in a predetermined "random" order, with different orders for each subject. The intervals were demarcated by the onset and termination of the appropriate lighting display.

Each subject was seated 2.5 m from the display, with eyes level to it. Wrist-

watches were removed prior to the experimental session. The following instructions were then read to each subject:

This is an experiment involving time perception. In front of you is a panel of lights which will go on for an interval of time. You will be asked to depress this switch [experimenter points to microswitch] to reproduce the duration of the interval. Beneath the panel of lights are 2 signal lights; the amber light on the left signals that the panel is about to come on, and the blue light on the right signals that you are to respond by depressing the switch. Please do not count or tap during the experiment. Do you have any questions?

Two practice intervals, 17 and 25 seconds, were presented to each subject prior to the experiment. The last 2 authors each ran half the subjects in every level of complexity.

RESULTS AND DISCUSSION

To compare directly the accuracy of estimation of different stimulus intervals, ratio transformations were performed on each estimation by every subject. The duration interval reproduced by each subject was divided by the measured time interval; thus ratio values of 1.00 show perfect accuracy, ratios below 1.00 indicate underestimation of the actual interval, and those above 1.00 indicate overestimation of the interval by a subject (see Hornstein and Rotter, 1969, for a further discussion of ratio transformations of time estimations).

A 3 × 3 × 6 analysis of variance was performed on the ratios, where the 3 factors were complexity, block, and duration, respectively. The main results are shown in Tables 1 and 2 and Figure 1, where mean

Table 1. Summary of anova for complexity, intervals, and blocks of trials

Source	Degrees of freedom	Sums of squares	Mean square	F
Complexity (A)	2	1.616	0 808	—
Subjects within groups (S)	60	54 739	0 912	
Block (B)	2	1 703	0.852	11.06[a]
AB	4	0 263	0.066	—
BS	120	9.214	0.077	
Interval (C)	5	1.892	0 378	6.87[b]
AC	10	1.197	0.120	2.18[a]
CS	300	16.423	0.055	
BC	10	0.832	0.083	2.59[b]
ABC	20	1.010	0.051	1.59[a]
BCS	600	19.204	0.032	

[a] $p < .05$
[b] $p < .01$

Table 2. Mean ratios by block, level of complexity, and stimulus interval

		Stimulus interval (sec)						
		15	19	23	27	31	35	$\bar{x}$
Block 1								
high		0.96	1.13	1.04	0.95	0.88	0.93	0.98
intermed.		0.99	0.92	1.12	1.01	0.96	0.95	0.99
low		0.94	0.94	0.86	0.95	0.90	0.96	0.93
	$\bar{x}$	0.96	1.00	1.01	0.97	0.91	0.95	
Block 2								
high		1.21	1.21	1.07	1.02	0.99	0.92	1.07
intermed.		1.03	0.97	1.02	0.98	1.03	0.94	1.00
low		1.09	0.96	0.92	0.96	0.96	0.89	0.96
	$\bar{x}$	1.11	1.05	1.00	0.99	0.99	0.92	
Block 3								
high		1.21	1.12	1.11	1.06	1.10	1.00	1.10
intermed.		1.20	1.11	1.04	1.11	1.08	0.96	1.08
low		1.03	1.00	0.99	1.06	0.95	0.95	1.00
	$\bar{x}$	1.15	1.08	1.05	1.08	1.04	0.97	

ratios are plotted by duration interval for 3 levels of stimulus complexity; for comparison of relative estimation error between the 2 stimulus series, the results of Schiffman and Bobko (1974) are also plotted.

Contextual effects

As illustrated in Figure 1, the shorter intervals are relatively overestimated and the longer intervals are underestimated; this main effect due to duration interval was significant at $F(5, 300) = 6.87, p < .01$, and lends support to Vierordt's Law for the stimulus intervals employed herein. The interaction of block by interval was also significant at $F(10, 600) = 2.59, p < .01$. As shown in Table 2, Vierordt's Law appears to be manifest only in blocks 2 and 3, i.e., contextual effects seem only to occur after the subjects have had an opportunity to establish the range of stimuli. Accordingly, a Newman-Keuls test was performed at each of the 3 blocks on the mean estimates per interval across levels of complexity (see column means in Table 2). At block 1, the mean estimate for 23 sec (i.e., 1.01 sec) differed significantly from the mean estimate for 31 sec (i.e., 0.91 sec) at $p < .01$; no other pairs of means differed significantly. At block 2, the mean estimate for 15 sec differed from the estimates for 23, 27, 31, and 35 sec, each at $p < .01$; the mean estimate for 19 sec and 35 sec also differed from each other at $p < .05$. At block 3, the mean estimate for 15 sec differed from that of 31 and 35 sec at $p < .01$, and from that of 23 sec at $p < .05$; also, the mean estimate for 35 sec differed from that

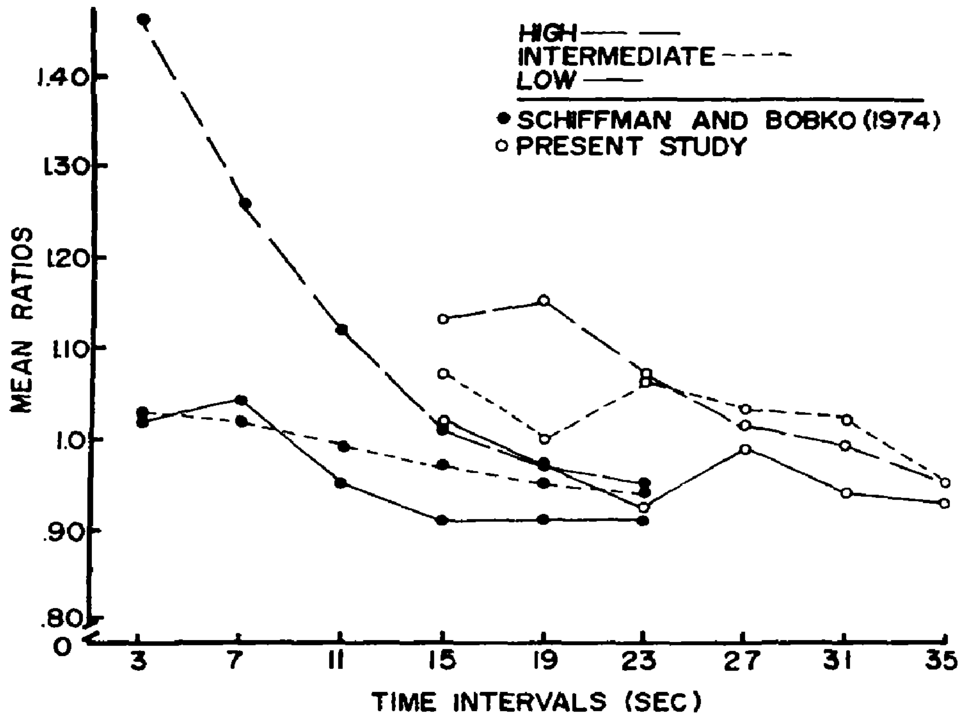


Figure 1. Mean ratios by duration interval for 3 levels of stimulus complexity (ordinate represents ratio of estimated duration of interval to actual interval)

of 27 sec at $p < .05$. Hence the relative overestimation of shorter intervals and underestimation of longer intervals that define Vierordt's Law tend to be manifest only after subjects become familiar with the range of the stimuli.

In order to compare the estimates for the stimulus intervals common to both the present experiment and the previous one (i.e., 15, 19, and 23 sec), the data were collapsed in each experiment across blocks for these three durations. A separate $2 \times 3 \times 3$ analysis of variance was performed on the mean ratios (i.e., the means of the 3 ratios per interval per subject); the factors were experiment (or context), complexity, and duration interval, respectively. As shown in Figure 1, the context of the stimulus intervals affected the duration estimates. Specifically, the stimulus intervals (physically identical in each experiment but contextually different) were relatively overestimated when they were the briefest members of the series but they were underestimated when they were the longest members of the series; this difference was significant at $F(1, 120) = 8.02, p < .01$. (The effects of stimulus complexity were not significant in this anal-

ysis. However, a significant main effect due to duration interval was found, $F(2, 240) = 3.66$, $p < .05$; estimates decreased with increases in stimulus duration. No interactions of the main effects were found to be significant.)

Practice effects

The central tendency effect predicts that duration estimates of shorter intervals increase and estimates of longer intervals decrease with practice or familiarity with the range of stimuli. As indicated in Table 2, however, such a trend is not apparent here; duration estimates tend to increase over blocks of trials for all stimulus intervals.

The main effect due to blocks was significant at $F(2, 120) = 11.06$, $p < .01$; the lengthening effect, i.e., the increase in estimates with practice, was more pronounced at the briefer durations (see Table 2). The interaction of block, interval, and complexity also reached significance at $F(20, 600) = 1.59$, $p < .05$. The interaction of block by complexity was not significant.

Thus the present findings agree with those reported by Eson and Kafka (1952), Falk and Bindra (1954), and Treisman (1963). (As Falk and Bindra note, the lengthening effect is apparently independent of any time-order errors; see also Woodrow (1935).) With the exception of a model by Treisman (1963), current theories of duration judgment neither include nor predict lengthening effects of the sort reported here and elsewhere. Briefly, Treisman's model requires a pulse generator whose rate may change in response to "specific temporal arousal," a counter that accumulates the number of pulses within a given stimulus presentation, a storage space, and a comparator. To account for a lengthening effect with the method of reproduction, it is assumed that the pulse-measure retrieved from storage and placed in the comparator for reproduction is some transformation of the original measure placed in storage (Postulate 6, pp. 20–21); that is to say, the retrieved measure is determined by the maximally activated location in the store, which in turn is a function of the last location to be activated *and* any residual effects of previous duration presentations. Hence, the reproduction of any given interval will be overestimated when the retrieved measure placed in the comparator is greater than the originally stored pulse-measure.

Complexity effects

Although Figure 1 indicates an increase in duration estimates with increases in stimulus complexity, a trend consistent with the earlier study, the main effects due to stimulus complexity were not found to be statisti-

cally significant. The interaction of complexity by interval, however, was significant at $F(10, 300) = 2.18, p < .05$; as indicated in Table 2 the effects of complexity are more pronounced at the briefer durations.

One possible explanation for this effect is that 2 different processes are involved in the estimation of duration intervals. Judgments of relatively brief temporal intervals are perhaps made with respect to the number, magnitude or complexity of the stimulus events defining and comprising the interval, while longer intervals are judged with respect to some other criterion or mechanism, e.g., rate of internal body rhythm. In other words, estimations of short intervals may be more affected by the characteristics of the external stimuli than are long intervals such that complexity has a diminishing effect as duration increases. Indeed, the results of Schiffman and Bobko (1974) appear due to the large effect of the high-complexity condition at *short* durations, and several investigators (Avant, Lyman, and Antes, 1975; Thomas and Weaver, 1975) have reported results consistent with this trend for tachistoscopic presentations of verbal stimuli varying in their familiarity. Also, Eisler (1975) has reported a break in the power function for duration experience, which may be taken to indicate differential processes in the perception of short and long intervals.

It should be noted that other investigators (e.g., Ornstein, 1969; Yeager, 1969) have reported significant effects due to stimulus complexity for durations comparable to those employed herein; however, methodological differences may contribute to the discrepancy between the present results and these earlier studies. For instance, Ornstein's procedure is basically one in which subjects are naïve with regard to the experimental task. That is, since subjects were not aware that their task was one of duration comparison until after the stimuli had been presented, it is more likely that judgments were made with respect to the information encoded about the stimulus characteristics rather than an encoding of duration *per se*. This reasoning follows a model of duration judgment discussed by Thomas and Weaver (1975) and Thomas and Brown (1974), where attention is shared during a stimulus presentation between a "timer processor" and an "information processor"; the encodings of both processors effect the estimate, but their relative influence depends on the "attention" devoted to each process. The model can also account for the diminished effect of complexity with increasing duration suggested in Figure 1; that is, as the duration of a stimulus increases, more attention can be given to the timer mechanism with a consequent increase in the reliability of estimates for longer intervals.

In conclusion, the data presented here clearly indicate the contextual nature of duration experience. Specifically, reproduction of a temporal

interval is affected by the position the interval occupies in the stimulus series. Thus, once the subject becomes familiar with the set of stimuli, the longer intervals in the series tend to be underestimated and the shorter intervals are overestimated. Although other studies have reported a contextual effect when duration estimates are made with a 7-point rating scale (for example, Braud and Holborn, 1966), the assimilation of judgments toward the mid-range value of a series evident in the present experiment occurs in the absence of any situationally relative or restricted response mode, e.g., category judgments. That is, contextual effects are evident with a rather direct and nonarbitrary judgmental indicator. In addition, a lengthening effect was manifest, and thus some predictions based on the central tendency of judgment were not met; the implication here is that regression and lengthening effects are independent of each other.

Notes

This research was supported by funds from the Research Council of Rutgers, The State University, Grant 07-2109, and the Charles and Johanna Busch Memorial Fund. Requests for offprints should be sent to H. R. Schiffman, Department of Psychology, Busch Campus, Rutgers, The State University, New Brunswick, NJ 08903. Received for publication March 17, 1976; revision, December 13, 1976.

References

Avant, L. L., Lyman, P. J., and Antes, J. R. 1975. Effects of stimulus familiarity on judged visual duration. *Perception and Psychophysics* 17:253–262.

Behar, I., and Bevan, W. 1961. The perceived duration of auditory and visual intervals: Cross-modal comparison and interaction. *American Journal of Psychology* 74:17–26.

Braud, W. G., and Holborn, S. W. 1966. Temporal context effects with two judgmental languages. *Psychonomic Science* 6:151–152.

Eisler, H. 1975. Subjective duration and psychophysics. *Psychological Review* 82:429–450.

Eson, M. E., and Kafka, J. S. 1952. Diagnostic implications of a study in time perception. *Journal of General Psychology* 46:169–183.

Falk, J. L., and Bindra, D. 1954. Judgment of time as a function of serial position and stress. *Journal of Experimental Psychology* 47:279–282.

Fraisse, P. 1963. *The psychology of time.* New York: Harper and Row.

Goldstone, S., Lhamon, W. T., and Boardman, W. K. 1957. The time sense: anchor effects and apparent duration. *Journal of Psychology* 44:143–153.

Helson, H. 1964. *Adaptation-level theory.* New York: Harper and Row.

Hollingworth, H. L. 1910. The central tendency of judgment. *Journal of Philosophical Psychology* 7:461–469.

Hornstein, A. D., and Rotter, G. S. 1969. Research methodology in temporal perception. *Journal of Experimental Psychology* 79:561–564.

Ornstein, R. E. 1969. *On the experience of time.* 1969. Baltimore, Md.: Penguin Books.

Parducci, A. 1974. Contextual effects: A range-frequency analysis. In *Handbook of perception,* vol. 2, ed. E. C. Carterette and M. P. Friedman. New York: Academic Press.

Postman, L., and Miller, G. A. 1945. Anchoring of temporal judgments. *American Journal of Psychology* 58:43–53.

Schiffman, H. R., and Bobko, D. J. 1974. Effects of stimulus complexity on the perception of brief temporal intervals. *Journal of Experimental Psychology* 103:156–159.

Schiffman, H. R., and Hamilton, L. W. 1974. A simple design for random light displays. *Behavior Research Methods and Instrumentation* 6:75–76.

Thomas, E. A. C., and Brown, I. 1974. Time perception and the filled duration illusion. *Perception and Psychophysics* 16:449–458.

Thomas, E. A. C., and Weaver, W. B. 1975. Cognitive processing and time perception. *Perception and Psychophysics* 17:363–367.

Treisman, M. 1963. Temporal discrimination and the indifference interval: implications for a model of the "internal clock." *Psychological Monographs* 77: (Whole No. 13).

Underwood, B. J. 1966. *Experimental Psychology.* New York: Appleton-Century-Crofts.

Woodrow, H. 1935. The effect of practice upon time-order errors in the comparison of temporal intervals. *Psychological Review* 42:127–152.

Woodrow, H. 1951. Time perception. In *Handbook of experimental psychology,* ed. S. S. Stevens. New York: Wiley.

Yeager, J. 1969. Absolute time estimates as a function of complexity and interruption of melodies. *Psychonomic Science* 15:177–178.

AMERICAN JOURNAL OF PSYCHOLOGY
December 1977, Vol. 90, No. 4, pp. 587–598

Long-term retention and the spacing effect in free-recall and frequency judgments

John J. Shaughnessy
Hope College

In 2 experiments undergraduate students were presented with a long list of words in which items were repeated varying numbers of times according to either a *massed presentation (MP) schedule or a distributed presentation (DP) schedule.* Following list presentation, subjects were given either a frequency-judgment test (Experiment I) or a free-recall test (Experiment II). In both experiments, the retention test was given either immediately after or 24 hours after list presentation. As expected, on the immediate test MP items were judged to have occurred less frequently and they were recalled more poorly than DP items. On the delayed frequency judgment test, the spacing effect was slightly (though not significantly) smaller, and these results were considered to be consistent with the prevailing notion that MP items are more poorly registered in memory. The results of the delayed free-recall test, however, were less consistent with this explanation in that there was no difference in the recall of twice-presented MP and DP items.

In a list presented for study, the successive occurrences of a repeated item may appear either in adjacent or nonadjacent list positions. Repeated items whose successive occurrences appear in adjacent positions are referred to as massed-presentation (MP) items, while those repeated in nonadjacent positions are referred to as distributed-presentation (DP) items. It has been consistently shown that, on tests of retention given immediately after presentation of the study list, MP items are more poorly recalled (e.g., Shaughnessy, Zimmerman, and Underwood, 1972) and are judged to have occurred less frequently (e.g., Underwood, 1969) than are DP items. Hintzman (1974) has provided a recent review of both the empirical findings and the proposed theoretical explanations related to this so-called spacing effect in free-recall and frequency judgments.

Somewhat surprisingly, very little attention has been given to the study of the spacing effect in long-term tests of retention. Ciccone and Brels-

ford (1976) did find a sizable spacing effect on a retention test given 24 hours after acquisition that involved a *continuous paired-associate* learning procedure. However, because no tests of retention comparable to those given after 24 hours were given immediately after the acquisition sequence, it is impossible to determine whether the size of the spacing effect was larger than, smaller than, or no different from that which would have been obtained immediately after acquisition. The purpose of the present experiments was to determine the relative size of the spacing effect in long-term tests of retention as compared to that obtained with tests of retention given immediately after presentation of the study list. Free-recall and frequency judgments were selected as the tests of retention because these have been the most commonly used measures in the research on the spacing effect.

EXPERIMENT I

METHOD

Design and subjects

Four independent groups of subjects were differentiated on the basis of the factorial combination of two independent variables. Half of the subjects were given a frequency-judgment test following list presentation and the other half were given a free-recall test. Of those subjects given each of the 2 types of tests, half were tested immediately after list presentation and half were tested 24 hours after list presentation. In addition, 8 forms of the study list were used in each of the 4 major conditions. Four undergraduate students were assigned to each of the 8 forms of the study list for a total of 32 subjects in each condition and 128 subjects in the experiment as a whole. To obtain the final total of 128 subjects, 134 subjects were run, but 5 of these failed to return for the 24-hour retention test and 1 failed to follow instructions for the immediate test.

Materials

Study list items were drawn randomly from a previously selected pool of 6-letter words with frequencies of 11-20 per million in the Kucera-Francis word count (1967). Seven words were chosen to be used as a primacy buffer with 6 of these words being presented once and one of the words being presented five times under MP. Four words each presented once served as the recency buffer. Four additional items served as filler items in the body of the list, with one each presented under DP at frequencies of 6, 7, 8, and 9. The items appearing as primacy, recency, and filler items were the same for all forms of the study list and data for these items will not be considered.

The body of the list consisted of 48 repeated items and 12 items presented once. Of the repeated items, half were presented under MP and half under DP. Of the MP items, 6 were presented twice, 6 three times, 6 four times, and 6 five times. The same was true for the DP items. Eight sets of 6 items each were constructed and these sets were randomly assigned to the 8 repeated-item conditions

to construct the first study-list form. Seven additional forms of the study list were then made by rotating the 8 sets of items such that across the 8 forms each set appeared in each of the 8 repeated item conditions once and only once. The same 12 items were used in all 8 forms for the once-presented items, but a different random assignment of items to specific list positions was used for each form. The body of the study list was divided into thirds with each third containing 4 once-presented items and 2 items from each of the 8 repeated-item conditions. The lags for DP items varied from 4 to 20 list positions. The position of last occurrence of a DP item of a given presentation frequency was yoked to that of an MP item of the same frequency. The entire study list consisted of 75 different words and 225 list positions.

The frequency judgment test booklet consisted of 6 pages of either 14 or 15 words printed in a single column with a blank next to each word. Each page contained 2 once-presented items; 1 item from each of the 8 repeated-item conditions; 2 zero-frequency items; and either 2 or 3 primacy, recency, and filler items. The 12 zero-frequency items were drawn from the same word pool as were the study-list items, and the same 12 items were used with each of the 8 study-list forms. All items were assigned randomly to pages of the test booklet and to positions on the pages of the test booklet.

Procedure

Subjects participated in the experiment in small groups of up to 4 students, and all subjects in a given group participated in the same condition. Groups of subjects were assigned to conditions according to a block randomized schedule. Prior to study-list presentation, all subjects were told that they would be presented with a long list of words and that they were to try to remember as many of the words as possible. They were told that the order in which the words were presented was unimportant and that some of the words would be repeated. The study list was then presented using a cassette tape recorder with the items presented approximately every 2.5 seconds.

Immediately following the presentation of the study list, those subjects assigned to the immediate free-recall condition were asked to write all the study-list words they could remember on the recall sheet provided. They were told that even if a word was presented more than once, they needed to write it on their recall sheet only once. Similarly, those subjects assigned to the immediate frequency judgment condition were told to assign to each word in the frequency-judgment booklet a number representing the number of times the word occurred on the study list. They were told to assign a value of zero to those words which had not appeared on the list, and they were told to do all the items on each page before going on to the next page. Those subjects in the delayed-test conditions were told immediately after the study-list presentation that they were not to be given the retention test until they returned at the same time the next day. They were instructed not to rehearse the study-list items and not to participate in other experiments before they returned for the retention test. When they returned the next day, they were given either the free-recall or frequency-judgment test using the same procedures as had been used for those subjects given the immediate tests.

RESULTS AND DISCUSSION

Frequency judgments for nonrepeated items

To determine the retention of frequency information for nonrepeated items, comparisons were made among several different procedures for scoring the frequency judgments given to the zero-frequency and once-presented items. The first scoring procedure involved simply determining the mean frequency judgment assigned by each subject to the zero-frequency and once-presented items. The means of these means for those subjects in the immediate and delayed conditions are presented in the upper-left portion of Table 1. These values are quite comparable to those reported by Underwood, Zimmerman, and Freund (1971) in their study of the long-term retention of frequency information. Overall, the mean judgments for once-presented items were reliably higher than those for zero-frequency items, $F = 94.84$, $MS_E = 0.37$.[1] Neither the overall increase in mean judged frequency over 24 hours nor the slight decrease over 24 hours in the difference between the mean judgments for once-presented and zero-frequency items was significant statistically ($F = 1.41$, $MS_E = 2.22$ and $F < 1, MS_E = 0.37$, respectively). These results might lead one to conclude that there was little or no loss over 24 hours in the ability to discriminate the frequencies of nonrepeated items. This conclusion is inconsistent, however, with those previously reported in the literature (e.g., Underwood, Zimmerman, and Freund, 1971) and, as

Table 1. Mean frequency judgments and recognition accuracy measures for nonrepeated items as a function of retention interval

	Frequency judgments			
	Raw scores		Transformed data	
	Imm.[a]	Del.[b]	Imm.	Del.
Zero-frequency items	.55	.91	.62	.86
Once-presented items	1.64	1.91	1.23	1.28
	Recognition accuracy measures			
	Mean number of recognition errors		Mean number of frequency judgment "hits"	
	Imm.	Del.	Imm.	Del.
Zero-frequency items	2.75	4.34	9.25	7.66
Once-presented items	3.25	3.91	4.03	2.97

[a] Immediate tests of retention.

[b] Delayed tests of retention given after 24 hours.

will be seen, with the conclusions based upon other analyses of the present data.

The next two analyses done involved the use of recognition accuracy measures derived from the frequency-judgment data. For the first of these analyses, the mean number of recognition errors for each subject for the zero-frequency and once-presented items was determined. Any non-zero frequency judgment given to a zero-frequency item was considered an error as was any judgment of zero given to a once-presented item. Underwood (1974) has shown that the sum of these 2 types of errors provides a "simple and meaningful measure of discriminability or sensitivity" (p. 921). The numbers of these recognition errors as a function of the retention interval variable are presented in the lower-left portion of Table 1. The only significant source of variation in the analysis of these means was the overall increase in errors from a mean total of 6.00 on the immediate test to one of 8.25 on the delayed test, $F = 15.76, MS_E = 2.57$.

The second measure of recognition accuracy was the number of frequency judgment "hits" for zero-frequency and once-presented items. This measure was introduced by Ghatala and Levin (1973) and it represents the number of judgments of zero given to zero-frequency items and the number of judgments of one given to once-presented items. This measure is obviously the complement of the recognition error measure for zero-frequency items but not for once-presented items. The mean numbers of frequency judgment "hits" for the 2 retention interval conditions are presented in the lower-right portion of Table 1. The judgments were more accurate for zero-frequency items than for once-presented items, $F = 162.54, MS_E = 4.83$; and, more critically, the judgments given immediately were more accurate than those given after 24 hours, $F = 7.27$, $MS_E = 7.76$. Both of the analyses of recognition accuracy provide evidence for the forgetting of frequency information for nonrepeated items over 24 hours.

The mean frequency judgments for once-presented items showed a slight, albeit nonsignificant, increase over 24 hours. This increase occurred even though there was an increase in the number of zero judgments given to once-presented items as evidenced by the analysis of recognition errors. This could occur only if higher judgments were given on the delayed test to those once-presented items which subjects judged to have occurred with some non-zero frequency. These overestimates (particularly large overestimates) would seem to carry undue weight in the analysis of the mean frequency judgment data. Therefore, an additional analysis was done using a square root transformation of the mean frequency judgments of nonrepeated items. The means of these transformed data are

presented in the upper-right portion of Table 1. As was the case in the analysis of the raw scores, the overall mean judgments for once-presented items were higher than those for zero-frequency items, $F = 169.00$, $MS_E = .05$. Contrary to the previous analysis, however, the analysis of the transformed data showed a statistically significant interaction of the frequency and retention interval variables, $F = 5.80$, $MS_E = .05$, indicating a loss over 24 hours in the ability to discriminate the frequencies of nonrepeated items.

Frequency judgments for repeated items

The mean judged frequencies for repeated items as a function of presentation frequency, MP-DP, and retention interval are presented in Figure 1. The mean judged frequency increased with increasing presentation frequency for both massed-presentation and distributed-presentation items in both retention intervals. The MP items were consistently judged to have occurred less frequently than the DP items of the same presentation frequency, and the size of the spacing effect tended to increase with increasing presentation frequency. The pattern of the change in the frequency judgments for DP items over 24 hours was similar to that previously reported in studies of the retention of frequency information (Leight, 1968; Underwood et al., 1971). That is, there was a tendency toward an increase in the judgments for low-frequency items and a decrease in the judgments for high-frequency items with the overall mean judged frequency holding relatively constant. This was not a pronounced change for the frequencies shown in Figure 1, but it must be remembered that this figure does not include the values for the highest presentation frequencies included in the study list. Finally, because of the consistent increase in the judgments for MP items over 24 hours at each presentation frequency, there was some suggestion of a reduction in the size of the spacing effect on the delayed test.

Analysis of the mean frequency judgments provided statistical support for the conclusions stated in the preceding paragraph. The overall mean judgments for presentation frequencies of 2, 3, 4, and 5 were 2.51, 2.77, 3.49, and 3.80, respectively; and this increase was reliable, $F = 33.47$, $MS_E = 1.39$.[2] Simple effects analyses indicated that the increase in judgments with increasing presentation frequency was statistically significant for each of the 4 curves in Figure 1. Overall, the mean judgment for DP items (3.60) was significantly higher than that for MP items (2.69), $F = 74.92$, $MS_E = 1.40$. Summing over the retention interval variable, the size of the spacing effect did increase with increasing presentation frequency, $F = 4.06$, $MS_E = 0.81$. The mean difference in the judgments

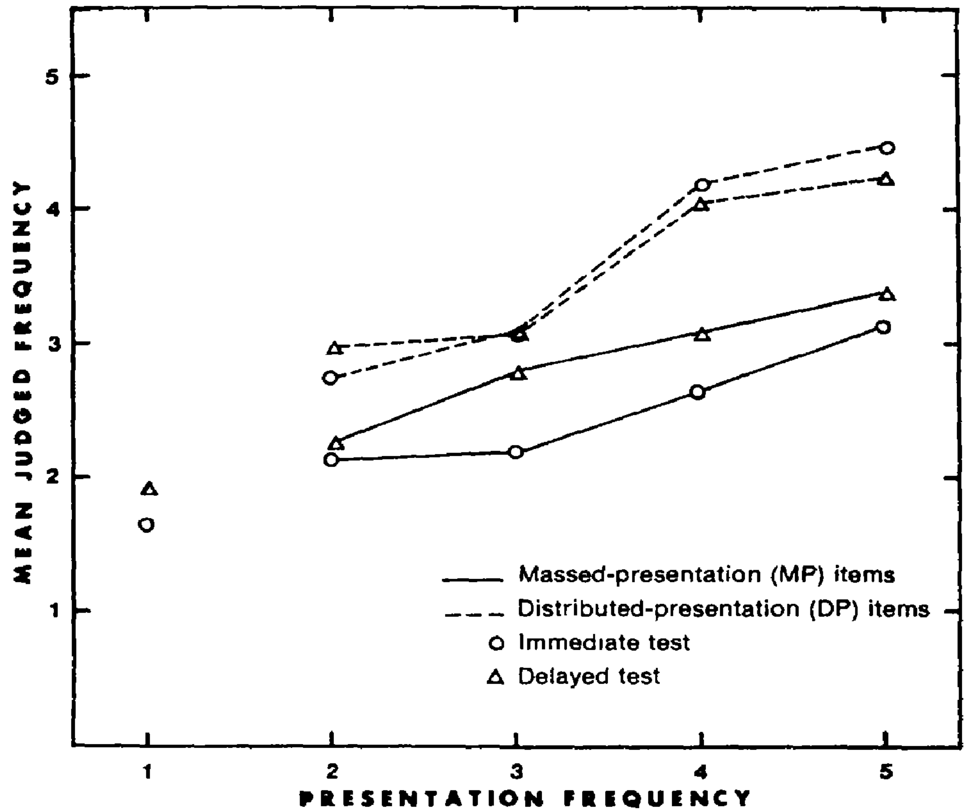


Figure 1. Mean judged frequency as a function of presentation frequency, MP-DP, and retention interval

for MP and DP items was slightly larger on the immediate test (1.10) than on the delayed test (0.71), but the interaction of the MP-DP and retention-interval variables fell just short of conventional levels for statistical significance, $F = 3.34$, $MS_E = 1.40$. Not too surprisingly, separate analyses showed the spacing effect to be highly reliable in both the immediate and delayed conditions.

Even though the change in the size of the spacing effect over 24 hours was small, it is tempting to try to provide a possible explanation for this outcome. This temptation is particularly strong given how difficult it has proven to be to find any independent variable that results in a decrease in the size of the spacing effect. Nonetheless, further examination of the present data argue against considering the change in the spacing effect over 24 hours as a reliable phenomenon in need of explanation. First, while the increase in frequency judgments over 24 hours for MP items

was consistent across presentation frequencies, the absolute magnitude of the increase, when tested separately, fell far short of statistical significance, $F < 1, MS_E = 43.96$. Further, because each of 48 different words appeared in each of the repeated item conditions equally often in the 2 retention interval conditions, it was possible to determine the number of times the mean judgment assigned to a given word serving as an MP item was higher after 24 hours than it was immediately. Of the 384 possible comparisons (48 words at 4 presentation frequencies), 53% resulted in a higher mean judgment for the word after 24 hours. The corresponding value for DP items was 48%. Finally, even if the increase for MP items were considered to be highly reliable, it would be expected if the low-frequency judgments given to MP items immediately were considered to be a true reflection of the deficient registration of these items. That is, one would expect items judged to have occurred less frequently than the mean list frequency on an immediate test to show an increase in judged frequency over 24 hours.

One final subsidiary analysis of the frequency-judgment data was suggested by a finding recently reported by Underwood, Kapelak, and Malmi (1976). They presented subjects with lists of 17 individual letters with some of the letters repeated 2 or 3 times under MP or DP. The retention test required subjects to identify the presented letters on a sheet containing an alphabetical listing of the 26 letters of the alphabet. One group of subjects was given the test immediately following presentation of the last study-list item; a second group was given the test after a 30-second delay. There was an overall spacing effect in both groups, but the effect was slightly smaller in the delayed condition. The finding of particular interest here, however, involved a breakdown of subjects in terms of their criterion for responding, as indexed by the total number of letters they circled on the retention test. Those subjects who circled more than the median number of letters were considered to have a lax criterion for responding, and the remaining subjects were considered to have a strict criterion. For those subjects having a strict criterion, the spacing effect was present on both the immediate and delayed tests. For those subjects having the lax criterion, the spacing effect was present on an immediate test, but MP items were more likely to be recognized than DP items on the delayed test.

The following analysis was done to determine if criterion differences could be shown to influence the size of the spacing effect when the retention test was given after a long rather than a short retention interval. First, a response-bias score was determined for each subject based upon the subject's judgments given to nonrepeated items. The bias measure

suggested by Underwood (1974) was used. Next, subjects in the immediate and in the delayed groups were divided at the median into a lax criterion and a strict criterion subgroup. Then, the size of the spacing effect for each of these 4 subgroups was determined. The mean difference between DP and MP items in the delayed condition for those subjects having a lax criterion was .71, while that for those with a strict criterion was also .71. The corresponding values in the immediate condition were 1.06 and 1.17. Needless to say, the criterion differences did not have the same effect in the present experiment as in the Underwood et al. (1976) study. The many differences between the two experiments, however, preclude specifying the reason for the widely different outcomes.

Free-recall results

Very little will be said about the recall data of Experiment I because of the low overall level of recall. Only 8% of the once-presented items were recalled in the immediate-recall condition. The mean percent recall for MP and DP items immediately was 14.25% and 25.25%, respectively; and the corresponding values in the delayed condition were 10.50% and 14.00%. The possible basement effect in the recall of MP items militates against an interpretation of the decrease in the size of the spacing effect over 24 hours. Therefore, a second experiment was done to examine the long-term retention of the spacing effect in free recall. To avoid the problem of a basement effect, the recall task was made considerably easier by changing the nature of the to-be-remembered material.

EXPERIMENT II

METHOD

Only two major groups of subjects were required for Experiment II, with one being given a free-recall test immediately after presentation of the study list and one being given the test 24 hours after list presentation. The 54 study-list items were high-frequency (A or AA in the Thorndike-Lorge *G* count) and high-concreteness (greater than 6.17 on the C-scale) nouns from the Paivio, Yuille, and Madigan (1968) norms. Six of the items served as a primacy buffer and 5 items were used for a recency buffer. One item was repeated 5 times under MP immediately following the primacy buffer. The same words served each of these functions in each form of the study list and recall for these items was not analyzed. Six words were presented once in the body of the list. The remaining words were divided into 6 sets of 6 words each and these words were used for the items repeated in the body of the list. Items were repeated

either 2, 4, or 6 times under either an MP or a DP schedule. One of the 6 sets was randomly assigned to one of the 6 repeated item conditions for the first form of the study list. Five additional forms of the study list were then constructed such that, across forms, each set was used for each repeated item condition once and only once.

The study list consisted of a total of 166 positions. The body of the study list was divided into thirds, with each third containing 50 positions. Two once-presented items and 2 of each of the 6 types of repeated items appeared in each third of the study list. The position of last occurrence of a DP item of a given frequency was yoked to that of the last occurrence of an MP item of the same frequency. Lags separating the repetitions of DP items ranged from 4 to 8 list positions. The same 6 items were used as once-presented items in each form of the study list, but the particular list position to which they were assigned was rotated across the 6 forms.

Five undergraduate students were assigned to each of the 6 study-list forms in each of the 2 retention intervals for a total of 60 subjects in the experiment. Three subjects failed to return for the 24-hour retention test and they were replaced by an additional 3 subjects. Subjects participated in the experiment in small groups of up to 5 subjects with all the members of a given group being assigned to the same condition according to a block-randomized schedule. The procedure and instructions were the same as those for the corresponding conditions in Experiment I, except that the study list was presented at approximately a 3-second rate.

RESULTS AND DISCUSSION

The mean number of words recalled as a function of presentation frequency, MP-DP, and retention interval is presented in Figure 2. The change in the to-be-remembered material was effective in raising the overall recall level as evidenced by the fact that recall of once-presented items after 24 hours in Experiment II was almost twice as good as the immediate recall of once-presented items in Experiment I. As would be expected, the mean total number of repeated items recalled immediately (16.57) was greater than that after 24 hours (11.90), $F(2, 96) = 24.52$, $MS_E = 1.14$. Summing over the 2 retention intervals and the 3 presentation frequencies there was a clear spacing effect with the mean total number of MP and DP items recalled being 12.27 and 16.20, respectively, $F(1, 48) = 33.63$, $MS_E = 1.15$. Although there was no overall interaction of the MP-DP and retention-interval variables ($F < 1$), there was some suggestion of a triple interaction involving these 2 variables and

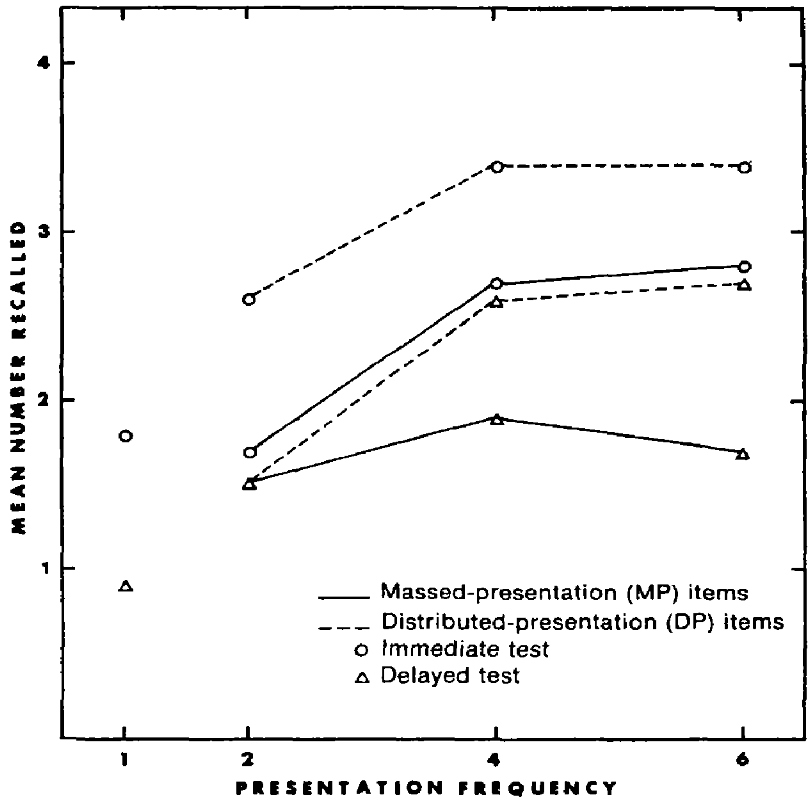


Figure 2. Mean number of words recalled as a function of presentation frequency, MP-DP, and retention interval (maximum possible recall of any one type of item being 6)

the presentation-frequency variable. The source of this 3-way interaction was obviously the absence of a difference in the 24-hour recall of MP and DP items presented twice. Subsequent analysis eliminated the possibility that this absence of a difference in recall was the result of a basement effect. The recall of MP-2 items in the 24-hour group was significantly higher than that of once-presented items, $F(1, 24) = 9.00$, $MS_E = .60$.

The absence of a spacing effect for twice-presented items after 24 hours is particularly interesting because so many of the studies examining the effects of the spacing of repetitions have involved a presentation frequency of 2. The prevailing notion that DP results in a more lasting and durable long-term memory trace seems consistent with the present results only for presentation frequencies of 4 and 6. At the very least, the delayed recall of twice-presented items in the present experiment can be added

to a growing list of findings (e.g., Underwood et al., 1976) suggesting that no single explanatory mechanism can be used to account for the various manifestations of the spacing effect.

Notes

I want to thank Eric Deaton, who was responsible for conducting these experiments and who assisted in the analysis of the results. Received for publication November 1, 1976.

1. The appropriate degrees of freedom for all *F*s reported for the analyses of the nonrepeated items were 1 and 48. Also, in all analyses to be reported in this paper, forms of the study list were included as a variable but effects involving this variable were not evaluated.

2. In the analysis of the frequency judgments for repeated items, the appropriate degrees of freedom for effects involving the presentation frequency variable were 3 and 144. For all other effects, the appropriate degrees of freedom were 1 and 48.

References

Ciccone, D. S., and Brelsford, J. W. 1976. Spacing repetitions in paired-associate learning: Experimenter versus subject control. *Journal of Experimental Psychology: Human Learning and Memory* 2:446–455.

Ghatala, E. S., and Levin, J. R. 1973. Developmental differences in frequency judgments of words and pictures. *Journal of Experimental Child Psychology* 16:495–507.

Hintzman, D. L. 1974. Theoretical implications of the spacing effect. In *Theories in cognitive psychology: The Loyola symposium*, ed. R. L. Solso. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Kucera, H., and Francis, W. H. 1967. *Computational analysis of present-day American English*. Providence, R.I.: Brown University Press.

Leight, K. L. 1968. Recall and judged frequency of implicitly occurring words. *Journal of Verbal Learning and Verbal Behavior* 7:918–923.

Paivio, A., Yuille, J. C., and Madigan, S. A. 1968. Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monograph Supplement* 76: No. 1, Part 2.

Shaughnessy, J. J., Zimmerman, J., and Underwood, B. J. 1972. Further evidence on the MP-DP effect in free-recall learning. *Journal of Verbal Learning and Verbal Behavior* 11:1–12.

Underwood, B. J. 1969. Some correlates of item repetition in free-recall learning. *Journal of Verbal Learning and Verbal Behavior* 8:83–94.

Underwood, B. J. 1974. The role of the association in recognition memory. *Journal of Experimental Psychology* 102:917–939. (Monograph)

Underwood, B. J., Kapelak, S. M., and Malmi, R. A. 1976. The spacing effect: Additions to the theoretical and empirical puzzles. *Memory and Cognition* 4:391–400.

Underwood, B. J., Zimmerman, J., and Freund, J. S. 1971. Retention of frequency information with observations on recognition and recall. *Journal of Exper*… [illegible] 140–162.

AMERICAN JOURNAL OF PSYCHOLOGY
December 1977, Vol. 90, No. 4, pp. 599–607

Expectancies as a determinant of interference phenomena

Lynn Hasher and Michael Greenberg
Temple University

One version, by Lockhart, Craik, and Jacoby, of a levels-of-processing model of memory asserts the importance of the role of expectancies about forthcoming information in determining the elaborateness of a memory trace: Expectancies that are subsequently confirmed are presumed to result in less-elaborated traces (via an abbreviation of the required number of cognitive operations) than expectancies that are disconfirmed. The present experiment was a test of the extension of this model to account for the buildup of and release from proactive interference seen in the Brown-Peterson task. The results of the experiment do not support this extension and particularly cast doubt upon the assumption that disconfirmations of expectancies result in especially elaborate memory traces.

The idea that special forms of processing are induced whenever there is a mismatch between an expectation and an outcome has been put forward to account for phenomena seen in various fields of psychology, including cognitive development (Kagan, 1970), social psychology (Brickman, 1972), information processing (Kahneman, 1973; Neisser, 1976) and animal and human learning (cf. Rescorla and Wagner, 1972; Rudy, 1974). Indeed, a recent revision (Lockhart, Craik and Jacoby, 1976) of the levels of processing model (Craik and Lockhart, 1972) also places some emphasis upon the importance of expectancies and their disconfirmations in human memory.

Lockhart, Craik, and Jacoby (1976) argue that expectancies are a potential source of processing efficiency because they provide a structure for incoming information that reduces the number of perceptual and "cognitive" operations required to encode a stimulus, at least *if* that stimulus confirms the expectancy. If however, the stimulus disconfirms the expectancy, more elaborate processing, or a greater number of operations, are required (cf. also Kahneman, 1973). It should be noted that within the framework of a levels-of-processing analysis of memory (cf.

Craik and Tulving, 1975), it is the elaborateness of the memory code that determines its accessibility. Unexpected events, with their elaborated memory traces should then be remembered better than expected events.

Among the various memory phenomena to which Lockhart et al. extend this aspect of their model is the buildup of interference seen in the Brown-Peterson task (cf. Lockhart et al., p. 99). The following is an elaborated version of their discussion. As the subject processes the initial triad of items at the semantic level, he or she encodes their categorical relation. This then becomes part of the expectancy the subject forms about the semantic aspects of items to be presented on future trials. As long as the category remains the same, this expectancy will be confirmed and the processing of items will be accomplished more efficiently than was the case on the first trial because of the reduction in the number of operations required to place the new item in a structure. This efficiency will, however, result in an increasingly sparse memory trace — an event reflected in the decline in performance across successive Brown-Peterson trials. One must also assume that there are a minimum number of cognitive operations required to match an incoming stimulus to an expectancy and consequently a maximally sparse memory trace to account for the fact that recall asymptotes after several trials in this task.

Such an explanation can easily be extended to account for the increased recall seen on category-change trials: A category change results in a mismatch between the expectancy and the outcome, requiring the subject to include the categorical identification stage in his analysis of the new triad. This more elaborate processing then results in a richer memory trace than would be expected were there no such mismatch. Note that this explanation of the buildup and release from interference does not depend upon the category change as its critical component; rather, it depends upon the number of cognitive operations required to assimilate the new information. Instances representing a new category require at least one, if not more, additional levels or stages of processing (the identification of the category itself) than do instances representing a category that has already been identified on the basis of previous experience and whose label had formed part of the expectancy device. This increase in the number of operations required to assimilate information will occur if the category is "new" to the subject in the context of the experiment or if the category is one that the subject is not expecting on a particular trial.

One should thus be able to vary the level of recall seen in the Brown-Peterson task by influencing the number of operations required to assimilate information. Indeed it should be possible to simulate "release," or an increase in recall, by increasing the number of operations required,

whether or not the category changes. One should also be able to maintain stable performance, by holding constant the number of operations required, again whether or not the category changes.

The present experiment is an explicit test of the extension of the Lockhart et al. expectancy hypothesis to the buildup and release of interference in the Brown-Peterson task. After a series of Brown-Peterson trials in which subjects' category expectancies were confirmed, a critical trial was introduced in which half the subjects had their category expectancies disconfirmed and half had them confirmed. This presumably required the former group of subjects to engage in at least one more operation — the discovery of the unexpected category — than was the case for the latter group of subjects. For half of the subjects in each of these two groups there was a category switch, for half there was not. If category expectancies determine the number of operations that in turn determine the richness of the trace, those subjects who receive a disconfirmation (and so have to add at least one operation) should show "release," or at least some increase in recall, whether or not the category changes, while those who receive a confirmation (and so may continue to do without a category abstraction stage at the time the triad is presented) should show less release, or perhaps even none, whether or not the category changes.

The present experiment was then a stringent test of the extension of the Lockhart et al. model to the Brown-Peterson task. In the obverse of the usual situation, a disconfirmation of a category expectancy was induced for subjects who continued in an old category and a confirmation was induced for subjects who received a new category. Comparing the performance of these conditions with that of the more typical "release" and "control" conditions enabled us to attribute release to the number of operations required by the confirmation or disconfirmation of an expectancy concerning the category membership of the forthcoming words, to the category-change operation or to both potential processes. The answer was quite clear: Category change was the sole determinant of performance.

In the present experiment the critical expectancies were induced by explicit instructions given to subjects concerning the category membership of the group of 3 words they were about to see. Thus in contrast to the usual Brown-Peterson task, subjects had 2 sources of information from which to form expectancies: experience with items on prior trials and the instruction. While it is conceivable that an instructionally induced expectancy is not the equivalent of an experientially determined one, it is unreasonable to assume that those provided by the experimenter have

no psychological consequences. Were this so, weather reports, look outs and other verbal warning systems would be useless in aiding the processing of information. In the psychological literature, there are verbal set effects that may be seen in both the perceptual (Neisser, 1976) and problem-solving literatures (Bourne, Ekstrand, and Dominowski, 1971). For example, a verbal instruction about the contents of a picture is as beneficial to processing speed as is having seen the actual picture (Potter, 1975). Nonetheless, various procedures were adopted to increase the validity of the experimenter-provided expectancy.

METHOD

Procedure

All subjects were tested on a series of ten 3-word Brown-Peterson trials, with the tenth trial, to be discussed later, serving as the critical experimental trial. The first 9 trials were divided into blocks of 3 successive trials whose instances were from the same taxonomic category. Thus category changes occurred on trials 4 and 7. The major change in procedure between the present study and others occurred in the introduction into the task of an overt, experimenter-induced category expectancy. This was effected by preceding each of the first 9 trials with correct information about the category membership of the succeeding instances.

One may assume on the basis of the Lockhart et al. model that a category name that precedes by several seconds a set of representative items will be processed both perceptually and cognitively, including, no doubt, on the semantic level. The likelihood of such processing was increased in the present instance by the use of familiar categories. This experimenter-provided expectancy should then serve in much the same way as the expectation subjects are presumed to derive on their own on the basis of a single trial in the standard version of the Brown-Peterson task. That is, it eliminates the need to process the category membership of subsequent representative instances. It should be noted here that Lockhart et al. suggest that expectancies can come from any of several sources, including immediate experience, previous experience, set, and instructions.

In order to increase the validity of the experimenter expectancy, and the likelihood that the subject will use it, two procedural decisions were made on the basis of pilot investigations: One was to allow the subject sufficient time for the category label to be assimilated; the other was to allow the subject sufficient experience with the confirmation of this expectancy (9 trials) prior to the introduction of a critical disconfirmation trial.

Each trial was comprised of the following series of events, the timing of each of which is shown in parentheses: an expectancy slide, which instructed the subject to expect instances of a particular category (5 sec); a blank slide (5 sec); the 3 category instances (2 sec); the distractor-task slide, containing a randomly chosen 3-digit number from which the subject counted backwards by three's (15 sec); a test slide, indicating to the subject that she or he should recall the category instances studied on that trial (15 sec). The subject was required to read the category instances aloud and to count aloud.

Prior to the first experimental trial, all subjects were fully informed of the details and timing of the procedure. They were then given a practice trial, paced at the experimental rate, with items from a category not used in the experiment proper. They also received practice in backward number counting. The materials were presented on slides by an externally timed projector. All subjects were tested individually.

The 4 critical experimental conditions occurred on the tenth trial where one-half of the subjects received a disconfirmation of their experimenter-induced expectancies. Those subjects who received a disconfirmation were divided into 2 groups differing in the way the disconfirmation was accomplished: For one subgroup of subjects it was effected by introducing items from a new category when the expectancy slide indicated they were to remain in the old category (disconfirmation + change); for the other subgroup of disconfirmed subjects it was effected by maintaining the previous category when the expectancy slide signaled a category change (disconfirmation + no change). Similarly, the subjects who received a confirmation of their expectancies were divided into 2 subgroups that differed in the source of the confirmation: One subgroup expected and received a category change (confirmation + change); the other expected and received no category change (confirmation + no change). The design of the critical trial can be conceived of as a 2 $\times$ 2 factorial combination of the category variable (change versus no change) with the expectancy variable (confirmation versus disconfirmation).

Materials

The materials were selected from an earlier study that used 12 triads from each of 4 categories: body parts, foods, clothing, and animals (Hasher, Goggin, and Riley, 1973). Four triads from each category were selected at random for use in the present study. Across all 4 experimental conditions the sequence of categories was counterbalanced such that each category served equally often in each of the 3 blocks comprising the first 9 trials. The 4 categories were also represented equally frequently on the tenth, critical trial. In addition, in order to make certain that performance on the critical trial was not determined by the unique triad presented, across all 4 conditions each of the 4 triads from a given category served in this position equally often. Within each condition any given category occurred on the final test trial for 12 subjects, with subgroups of 3 subjects receiving the identical triad. This assignment procedure required 48 subjects per condition. For all other trials the particular sequence of triads within a category was randomly determined.

Subjects

The subjects were Temple University undergraduates who received course credit for their participation. Several subjects were initially discarded because of mechanical failures while others were because they failed to follow the counting directions. A few subjects had to be discarded because they either could not or did not read the 3 stimulus items in the 2-sec presentation interval. After the data collection phase was completed, with 48 subjects in each of the 4 conditions, we discovered that one subject had to be discarded because of an experimenter error on the tenth, critical trial. Thus there were 48 subjects in 3 of the conditions but only 47 subjects in the disconfirmation + no change condition.

RESULTS AND DISCUSSION

The dependent measure was the number of items, out of three, recalled on a given test trial. The means for each condition on the first 9 test trials may be seen in Table 1. All significant results are so at or beyond the .05 level.

Comparability of groups

On the first 9 test trials, all 4 groups of subjects were treated identically, and it was important that there be no performance differences among groups on these precritical trials. That this was the case, with the possible exception of trial 9 is clear from an inspection of Table 1. Despite appearances, however, there were no differences among the groups on trial 9, $F(3, 187) = 2.36$. The trial 9 data were also analyzed by sorting the groups into the treatment conditions they were to receive on trial 10. This 2 (confirmation versus disconfirmation) by 2 (category change versus no change) analysis revealed effects of borderline significance, at the .05 level, for the dummy confirmation manipulation, $F(1, 187) = 3.24$, $MS_e = .85$, and for the dummy category change manipulation, $F(1, 187) = 3.87$. Although all groups were treated identically on trial 9, the 2 groups that were to have their expectancies confirmed on trial 10 performed slightly better on trial 9 (2.18) than did the 2 groups that were to have theirs disconfirmed (1.94). In addition, the 2 groups that were to continue to receive the same category on trial 10 recalled more words on trial 9 (2.19) than did those groups that were to receive a new category (1.93). Why this pattern should appear on trial 9, given that all subjects were treated alike on this and on all preceding trials and also given that performance differences prior to trial 9 were minimal, is not obvious. These differences do not, however, invalidate the comparisons of interest on trial 10, although they must be considered in any interpretation of those results.

Table 1. Mean number recalled in first nine test trials

	Test Trial Number								
Condition	1	2	3	4	5	6	7	8	9
Expectancy confirmed									
Category change	2.65	2.52	1.98	2.71	2 29	1.92	2.52	2.19	2.04
No change	2.69	2 21	1.94	2.60	2.29	2.10	2.58	2.04	2.31
Expectancy disconfirmed									
Category change	2.52	2.15	1.94	2.56	2.23	2 00	2.54	2.19	1.81
No change	2.53	2.38	1.91	2.68	2.19	1.98	2.62	2.30	2.06
Mean	2.60	2.31	1.94	2.64	2.25	2.00	2.57	2.18	2.06

Replication of PI buildup and release phenomena

The buildup of interference as subjects continue to receive instances from the same category and the dissipation of that interference as the category changes are well-documented experimental findings (cf. Wickens, 1970). Our data present no exception, as can be seen in Table 1, with all 4 conditions showing a significant buildup of interference on trials 1–3, 4–6, and 7–9 and also significant release effects on trials 4 and 7. There are 2 noteworthy aspects of these results: For one, all subjects on the category-change trials (4 and 7) had been warned, via the expectancy slide, that instances from a new category would be forthcoming. When these instances appeared, 5 sec later, they theoretically should have confirmed the subjects' experimentally induced expectancies. These were both situations in which no disconfirmation occurred, yet recall improved. Second, the extent of release did not diminish much from the first category change trial (4) to the second (7) even though on this latter occasion subjects had more experience with the validity of the expectancy slide.

Critical trial performance

Performance on trial 10 was analyzed using a 2 (confirmation versus disconfirmation) by 2 (category change versus no change) analysis of variance. This analysis, easily confirmed by an inspection of Figure 1,

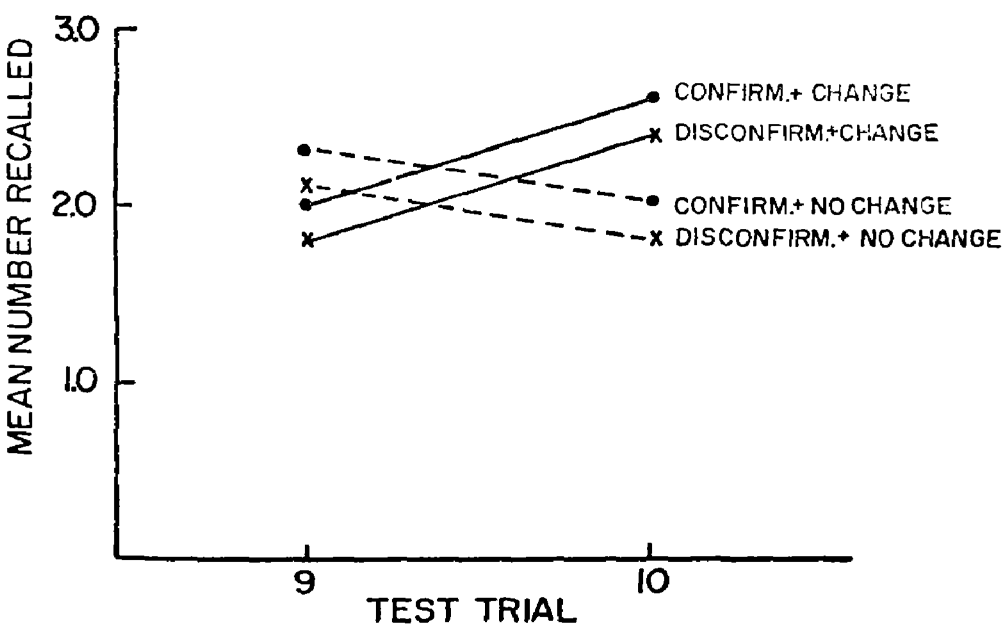


Figure 1. Mean number of words recalled on trials 9 and 10

shows clearly that performance of the 2 conditions that received a category change was superior to that of the 2 conditions that received no change, $F(1, 187) = 22.80, MS_e = .82$. Whether a subject's expectancy was confirmed or disconfirmed had no significant effect on performance, $F(1, 187) = 2.49$. While the mean recall (2.12) for the disconfirmed groups was lower than the mean recall (2.32) for the confirmed groups, a trend in the same direction was also present on trial 9, prior to the introduction of the discrepancy manipulation. Finally the expectancy procedure did not interact with the category procedure, $F < 1$.

A comparison of performance on trial 9 with that on trial 10 (Figure 1) also shows clearly that improved recall is associated only with the introduction of a category change. Expectancies play no significant role in influencing performance. An analysis of variance that included trials as an additional factor confirmed these conclusions.

It is clear from these results that release from interference in the Brown-Peterson task cannot be attributed to a presumed disconfirmation of the expectancy that the new instances will be members of the old category; subjects whose expectancies for a new category were violated by the maintenance of an old category did not show "release." It is also clear that the continued presence of interference for the typical control subjects in this task — subjects who continue to receive instances of the old category — is not induced simply by minimal processing that is the result of the confirmation of a category expectation; subjects whose category expectations were confirmed with instances from a new category did in fact show "release."

With regard to the Lockhart, Craik, and Jacoby model, one must either conclude that recall is not a function of the elaborateness of the trace (but see Craik and Tulving, 1975), or that disconfirmations of expectancies do not produce more elaborate traces than do confirmations of expectancies, at least when those expectancies concern the categorical membership of new groups of words in the Brown-Peterson task.

Notes

This research was supported by a grant (MH27755) to the first author from the National Institute of Mental Health. We would like to thank Walter Chromiak for his help with the pilot stages of this research and Joyce Iavecchia for her help with the final stages of the research. Robert Weisberg also made substantial contributions to the project. Received for publication July 26, 1976; revision, December 28, 1976.

References

Bourne, L. E., Jr., Ekstrand, B. R., and Dominowski, R. L. 1971. *The psychology of thinking*. Englewood Cliffs, N.J.: Prentice-Hall.

Brickman, P. 1972. Rational and nonrational elements in reactions to disconfirmation of performance expectancies. *Journal of Experimental Social Psychology* 8:112–123.

Carey, S. T. 1973. Delayed recognition testing, incidental learning, and proactive-inhibition release. *Journal of Experimental Psychology* 100:361–367.

Craik, F. I. M., and Lockhart, R. S. 1972. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior* 11:671–684.

Craik, F. I. M., and Tulving, E. 1975. Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General* 104:268–294.

Hasher, L., Goggin, J., and Riley, D. A. 1973. Learning and interference effects in short-term memory. *Journal of Experimental Psychology* 101:1–9.

Kagan, J. 1970. Attention and psychological change in the young child. *Science* 170:826–832.

Kahneman, D. 1973. *Attention and effort*. New York: Prentice-Hall.

Kintsch, W. 1970. *Learning, memory, and conceptual processes*. New York: Wiley.

Lockhart, R. S., Craik, F. I. M., and Jacoby, L. 1976. Depth of processing, recognition, and recall. In *Recall and recognition*, ed. J. Brown. London: Wiley.

MacLeod, C. M. 1975. Release from proactive interference: Insufficiency of an attentional account. *American Journal of Psychology* 88:459–465.

Neisser, U. 1976. *Cognition and reality: Principles and implications of cognitive psychology*. San Francisco: Freeman.

Potter, M. C. 1975. Meaning in visual search. *Science* 187:965–966.

Rescorla, R. A., and Wagner, A. R. 1972. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In *Classical conditioning II: Current theory and research*, ed. A. H. Black and W. F. Prokasy. New York: Appleton-Century-Crofts.

Rudy, J. W. 1974. Stimulus selection in animal conditioning and paired-associate learning: Variations in the associative process. *Journal of Verbal Learning and Verbal Behavior* 13:282–296.

Schendel, J. D. 1976. Analysis of the rehearsal processes in the release from proactive interference. *Journal of Experimental Psychology: Human Learning and Memory* 2:76–82.

Wickens, D. D. 1970. Encoding categories of words: An empirical approach to meaning. *Psychological Review* 77:1–15.

continued from page 576

Platt, Jerome J., and Labate, Christina. *Heroin Addiction.* New York: Wiley, 1976. Pp. 417. $19.95.

Poduska, Bernard. *You Can Cope: Be the Person You Want to be through Self-Therapy.* Englewood Cliffs, N.J.: Prentice Hall, 1976. Pp. 209. $7.95.

Pollak, Otto. *Human Behavior and the Helping Professions.* New York: Halsted, 1976. Pp. 98. $7.95.

Polland, Barbara K. *Decisions, Decisions, Decisions.* Millbrae, Calif.: Celestial Arts, 1976. Pp. 109. Paperback. $4.95.

Poole, Millicent E., and others. *Before School Begins.* New York: Wiley, 1976. Pp. 345. Paperback. $7.95.

Powell, John. *Fully Human Fully Alive.* Niles, Ill.: Argus Communications, 1976. Pp. 189. Paperback. $1.95.

Power, P. G., ed. *Adolescent Development.* Melbourne: Macmillan, 1976. Pp. 372. Paperback.

The Psychology Calendar. Northampton, Mass.: Mulch Press, 1976. Paperback. $4.95.

Rassieur, Charles L. *The Problem Clergymen Don't Talk About.* Philadelphia: Westminster Press, 1976. Pp. 157. Paperback. $3.95.

Raven, Bertram H., and Rubin, Jeffrey Z. *Social Psychology: People in Groups.* New York: Wiley, 1976. Pp. 591. $12.95.

Readings from Scientific American: Progress in Psychobiology. San Francisco: W. H. Freeman, 1976. Pp. 392. $14.00.

Reed, David. *Anna.* New York: Basic Books, 1976. Pp. 243. $10.00.

Reuchlin, Maurice, ed. *Cultures et Conduites.* Paris: Presses Universitaires de France, 1976. Pp. 356. Paperback.

Rexford, E. N.; Sander, L. W.; and Shapiro, T., eds. *Infant Psychiatry: A New Synthesis.* New Haven, Conn.: Yale University Press, 1976. Pp. 350. $20.00.

Reynolds, David K., and Farberow, Norman L. *Suicide: Inside and Out.* Berkeley, Calif.: University of California Press, 1976. Pp. 232. $10.95.

Ribes-Inesta, Emilio, and Bandura, Albert, eds. *Analysis of Delinquency and Aggression.* New York: Halsted, 1976. Pp. 237.

Riegel, Lakus, ed. *The Development of Dialectical Operations.* Basel: S. Karger, 1976. Pp. 241. Paperback. $15.00.

Riegel, Klaus F., and Meacham, John A. *The Developing Individual in a Changing World.* Vols. I and II. Chicago: Aldine, 1976. Pp. 736. $43.00.

Riesen, Auston H., and Thompson, Richard F., eds. *Advances in Psycholobiology.* Vol. Three. New York: Wiley, 1976. Pp. 489. $24.95.

Ritvo, Edward R., ed. *Autism.* New York: Halsted, 1976. Pp. 302. $16.95.

Roberts, Joan I., ed. *Beyond Intellectual Sexism.* New York: David McKay, 1976. Pp. 386. Paperback. $7.95.

Robinson, P. W. *Fundamentals of Experimental Psychology: A Comparative Approach.* Englewood Cliffs, N.J.: Prentice-Hall, 1976. Pp. 384. $12.95.

Rodgon, Maris Monitz. *Single-word Usage, Cognitive Development and the Beginnings of Combinatorial Speech.* New York: Cambridge University Press, 1976. Pp. 163. $14.95.

continued on page 634

AMERICAN JOURNAL OF PSYCHOLOGY
December 1977, Vol. 90, No. 4, pp. 609–620

The case for misapplied constancy scaling: Depth associations elicited by illusion configurations

Lawrence M. Ward, Clare Porac, Stanley Coren, and Joan S. Girgus
University of British Columbia, University of Victoria,
University of British Columbia, and City College, CUNY

It has been suggested that many visual-geometric illusions arise from inappropriate evocation of size-constancy by depth cues implicit in illusion configurations. Observers gave free association responses while viewing illusion figures. Analysis of these responses provides weak but consistent evidence for the elicitation of depth in the Sander parallelogram, Mueller-Lyer, Zoellner, and Ehrenfels variant of the Ponzo illusion. No evidence for depth is found in the normal form of the Ponzo, Poggendorff, and horizontal-vertical illusions, and the evidence is ambiguous in the Orbison configurations. These results indicate that depth processing may be evoked by some, but not all, classical illusion forms.

In the early 1960s Gregory (1963) resuscitated an old theory of visual illusions first proposed by Thiery (1896). This theory suggests that illusion configurations contain pictorial depth cues that may prompt observers to interpret the configuration as a 2-dimensional representation of a 3-dimensional array. If these implicit depth cues are sufficiently strong, they could evoke the constancy-scaling mechanisms for size or shape. However, the constancy scaling that results, while appropriate for a real 3-dimensional scene, is inappropriate for the 2-dimensional configuration. The theory proposes that, as a result of this misapplication of the constancy-scaling mechanism, the final percept contains distortions in the apparent shape or size of various elements.

Depth processing, and the resultant inappropriate constancy scaling, have been used to explain a number of visual distortions including the Mueller-Lyer illusion (Gregory, 1963, 1968), the Ponzo illusion (Kilbride and Leibowitz, 1975; Leibowitz, Brislin, Perlmutter, and Hennessey, 1969; Leibowitz and Pick, 1972), the Poggendorff illusion (Gillam,

1971; Green and Hoyle, 1963), the horizontal-vertical illusion (Girgus and Coren, 1975), several relative-size illusions such as the Baldwin configuration (Day, 1972), and the overestimation of width within open curves (Coren and Festinger, 1967), as well as some newly created illusions (Coren and Girgus, 1975).

The most powerful evidence for depth-cue involvement in visual illusions comes from some direct measurements made by Gregory (1966, 1968). He used a dimly illuminated Mueller-Lyer figure, presented monocularly against a dark background. This technique eliminated disparity and textural cues that may provide information about the actual distance of parts of the configuration. Observers were required to adjust a binocularly seen point of light so that it appeared to be at the same distance as various points on the illusion configuration. The data indicate that, in the apparently longer half of that figure, the wing tips are seen as closer to the observer than the shaft, while in the apparently shorter half, the shaft appears closer than the wing tips. These results are consistent with the theory that the converging wings serve as perspective cues which cause the configurations to be seen in depth, thus triggering constancy-scaling mechanisms. Coren and Festinger (1967) utilized a similar technique and obtained similar results on a curved variant of the Ponzo illusion.

To the extent that there are depth cues in illusion configurations, one would expect appropriate relative distance relationships to appear in observers' reports about such stimuli. Unfortunately, one seldom obtains spontaneous reports of this nature, and even prompting observers has not produced consistent results (Hotopf, 1965; Pike and Stacey, 1968; Worrall, 1974). Of course, it is possible that the textural cues and information from binocular disparity, which indicate that the configuration is flat and which were eliminated in the experiments by Gregory (1966, 1968) and Coren and Festinger (1967), may, under the usual presentation conditions, be sufficiently strong to offset the spontaneous appearance of depth. Nonetheless, it seems reasonable to assume that if implicit depth cues are present in the stimulus array, and if observers view the stimulus as a pictorial representation, any reports of relative depth ought to be consistent with the directions needed to produce the illusory distortion. In addition, the relative frequency of reports of phenomenal depth in illusion configurations obtained from a large sample of observers should serve as an index of the likelihood that depth cue processing is involved in a particular array. The following experiment was conducted to test these hypotheses.

METHOD

Subjects

Subjects were 120 college-student volunteers.

Stimuli

There were 13 different stimulus configurations: 6 were standard illusion figures (Figure 1), 5 presented only the inducing lines of standard illusion figures (Figure 2), and 2 presented a reduced version of a standard illusion configuration (Figure 3).

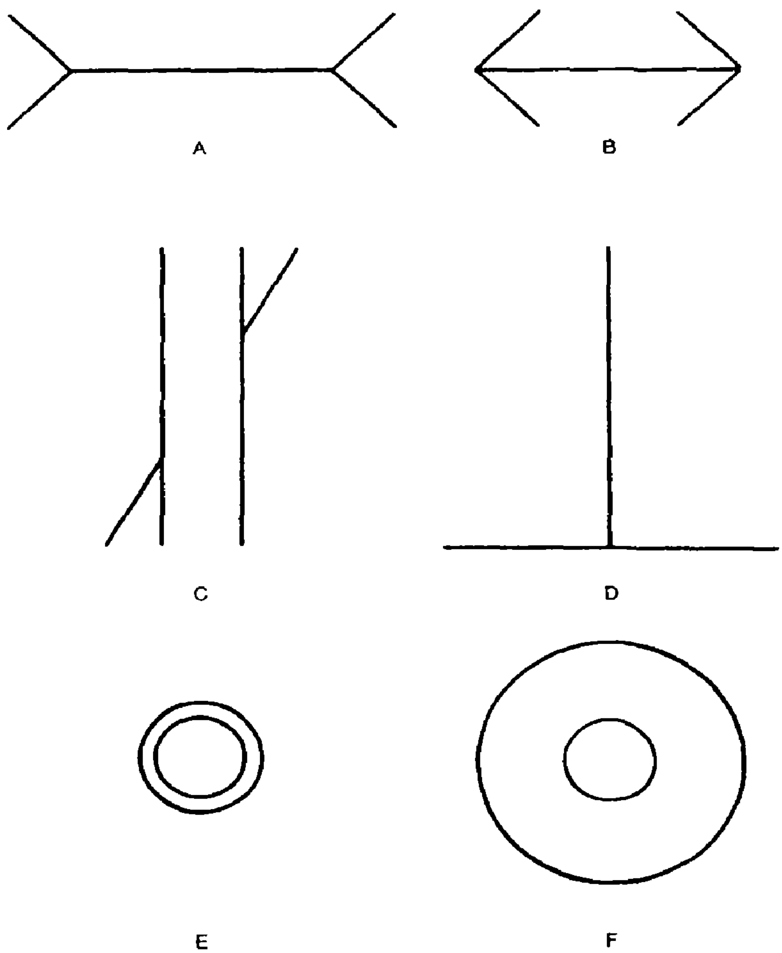


Figure 1. Classical illusion configurations used in this experiment: A shows the apparently longer half of the Mueller-Lyer illusion while B shows the apparently shorter half; C shows the Poggendorff illusion; D shows the horizontal-vertical illusion; E shows the apparently larger half of the Delboeuf illusion while F shows the apparently smaller half

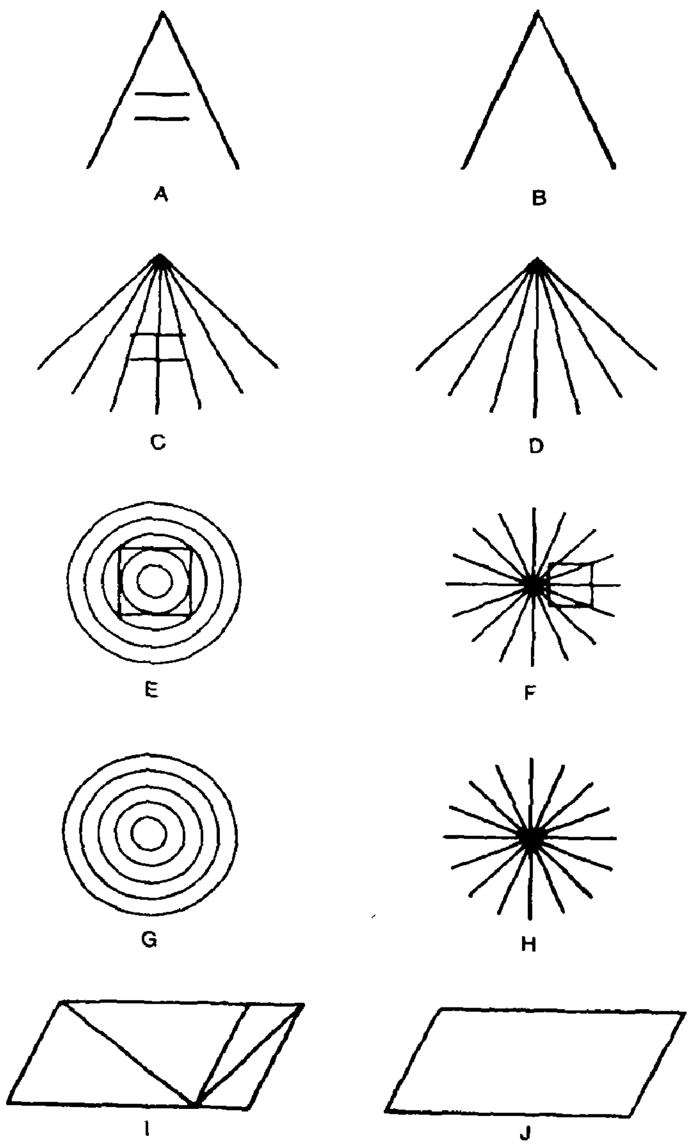


Figure 2. Inducing components of illusion configurations used in this experiment: A shows the usual form of the Ponzo illusion while B shows the inducing elements that were presented; C shows the Ehrenfels version of the Ponzo illusion and D shows the inducing elements for this configuration; E and F show two standard Orbison configurations while G and H show the inducing elements which were used; I shows the usual version of the Sander parallelogram; and J shows just the inducing lines of this figure

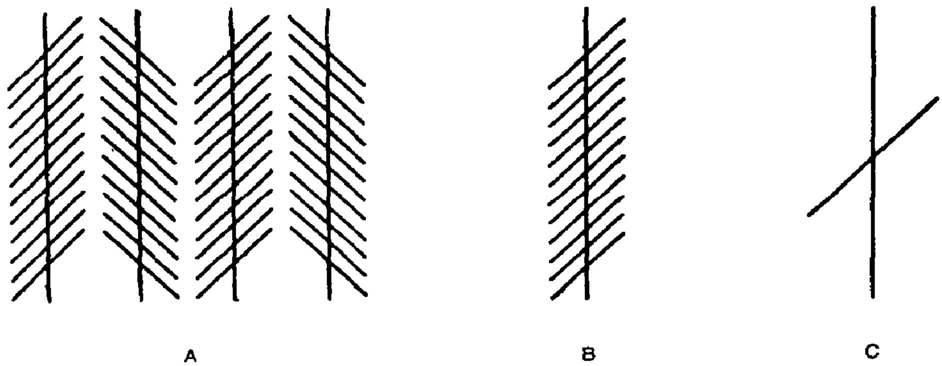


Figure 3. Reduced illusion configurations used in this experiment: A shows the usual form of the Zoellner illusion; B shows the reduced version employed; and C shows a minimal version of this illusion that was also used

Procedure

The illusion components and inducing segments were printed one to a page, in lines 1 mm wide. Each subject received the 13 stimulus configurations in a different random order. Subjects were instructed as follows: "Every set of lines can be viewed as a primitive picture of an object or a scene. On each page you will find a set of lines making up some sort of array. Imagine that these lines were drawn by a child attempting to depict an object or scene. What object or scene was he trying to represent?" Observers were required to write on the lower half of each page a description of the object or scene that they thought was represented by the stimulus configuration presented on that page.

RESULTS AND DISCUSSION

Three judges independently rated the depth relationships indicated in each subject's response to each stimulus. Blanks and responses that did not produce full agreement among the 3 judges were excluded from the data analysis. Such responses accounted for less than 2% of the data.

Table 1 provides a summary of a typical response breakdown, using shortened response categories and the responses emitted for Figure 1A. Classification of a given response was based either on the subject's stated point of view (as in "birds eye view of a roof"), or on the judge's assumption that the most normal viewing angle had determined the depth encoding.

Response length was of course not controlled. In order to determine whether the probability of a depth response increased with response length, a check was performed comparing the lengths (number of words) of the responses listed in Table 1 as depth and as non-depth reports. A

Table 1. Frequency distribution of scored responses to apparently longer half of Mueller-Lyer figure (Figure 1A)

Shaft near		Shaft far		Ambiguous		No depth	
Houseroof	(5)	Corner of room	(8)	Seesaw	(1)	Stickman	(31)
Bench	(4)	Book	(5)			Coat rack	(13)
Sawhorse	(3)	Bathtub	(1)			Stand	(6)
Spit	(2)	Balance beam	(1)			Arrow	(4)
Hurdle	(2)	Inverted hurdle	(1)			Cellwall	(3)
Dog	(2)	Tunnel	(1)			Dog bone	(2)
Bridge	(1)					Glass	(2)
Swingset	(1)					Dog	(1)
Badminton net	(1)					Birdbeaks and worm	(1)
						The "rack"	(1)
						Funnel	(1)
						Tree	(1)
						Road	(1)
						Happiness	(1)
						Lines	(1)
						Weights	(1)
						Umbrella	(1)
						Antenna	(1)
						Something tearing	(1)

median split produced a χ^2 of only 0.21, which seems to indicate no such relationship.

Let us now turn to the actual pattern of responses. In order to simplify our discussion, we will deal with each illusion configuration in turn.

Mueller-Lyer illusion (Figures 1A and 1B). The constancy-scaling explanation of the Mueller-Lyer configuration requires that the wings serve as perspective cues that indicate that the shaft in the apparently longer portion is more distant than the wings; for the apparently shorter component, the wings should serve as perspective cues that indicate that the shaft is closer. First, the number of depth responses to these configurations was not exceedingly large. For the apparently longer segment (Figure 1A), 34% of the responses describe objects or scenes that imply relative depth between the wings and the shaft, while 46% of the responses imply phenomenal depth in the apparently shorter segment (Figure 1B). For the apparently shorter segment, all of the depth responses are in the expected direction (obviously different from the proportion expected under the hypothesis of "chance"). They imply that the shaft is closer than the wing tips, as in "a roof viewed from above." However, 53% of

the depth responses to the apparently longer half of the Mueller-Lyer illusion imply that the shaft appears closer than the wing tips. These phenomenal reports are in opposition to a prediction from a constancy-scaling hypothesis. Nonetheless 47% of responses to the apparently longer half, which imply depth, are consistent with the predicted relationship (e.g., "a sawhorse turned on its back"). This proportion (.47) is not different from that expected under the "chance" hypothesis ($z = .33$). Perhaps there is some configurational tendency in the Mueller-Lyer figure that produces a constant error that makes the shaft look closer. This tendency would work in conjunction with the implicit depth cues in the apparently shorter segment of the illusion. However, it would work against the implicit depth cues in the apparently longer segment leading to the pattern of depth responses reported here.

The Poggendorff illusion (Figure 1C). Gillam (1971) has suggested a depth-processing interpretation for the Poggendorff illusion that depends upon the parallel lines being seen as a plane facing the observer frontally. The ends of the transversal intersecting with the parallels are seen as being co-planar but receding into the distance. Of all responses to this stimulus, 56% describe a scene or object with consistent depth. All of these reports indicate that the parallel lines are seen as representing a frontal surface that is parallel to the observer. However, none of these reports indicate that the transversal is seen as Gillam has proposed. Rather the transversal is seen as passing behind or through the frontal surface as in "a bridge over a railway track" or "a spear stuck through a tree." It is interesting to note that, despite the fact that the illusion measures close to a centimeter in this configuration, the descriptions of the transversal always imply that it is seen as continuous.

The horizontal-vertical illusion (Figure 1D). Girgus and Coren (1975) have argued that the horizontal-vertical illusion can be explained in terms of misapplied constancy scaling. They suggested that the horizontal line is always seen as lying parallel to the frontal plane, while the vertical line may be seen as slanted backwards into space. When we look at the phenomenal responses to the horizontal-vertical illusion, however, we find that only 21% of the responses imply any depth whatsoever. Of these, only one implies that the vertical is seen as slanted backwards into space. However, of those responses that do imply relative depth, 76% indicate that the vertical line is seen as being farther away than the horizontal (e.g., "floor-level view looking into a corner"). Such reports could lead to an explanation of the illusion in terms of constancy scaling, but the small number of depth responses argues against such an interpretation.

The Ponzo illusion (Figures 2A, 2B, 2C and 2D). The Ponzo illusion is

frequently interpreted as a clear example of constancy scaling based upon the converging lines that serve as pictorial perspective cues (Leibowitz and Pick, 1972). Since, when it is pointed out, it is easy to see the converging line elements in Figure 2B as the converging sides of a road or railroad track, it is surprising that only 3% of the entire sample gave such relative depth responses. By far the most prevalent response (55%) is that of an inverted cone, as in a "mountain" or "teepee." Such responses do not seem to imply perspective depth. However, they could be considered to be indications of weak phenomenal depth, since the base of such objects would be nearer to the observer than the apex. Still the total depth response to this configuration must be considered to be minimal.

When one looks at the inducing portion of the Ehrenfels form of the Ponzo (Figure 2D), however, the picture changes dramatically. For this configuration, 74% of the sample gave consistent depth responses, and all of these indicate that the apex is the most distant part of the array, as in "a multilaned highway going off into the distance." These data imply that phenomenal perspective depth is not generally spontaneously apparent in the normal Ponzo configuration. However, by increasing the number of inducing elements, relative depth, as predicted by constancy scaling theory, is readily elicited.

The Zoellner illusion (Figures 3A, 3B, and 3C). In order to predict the Zoellner illusion based upon implicit depth cues, it must be argued that the slanted lines are seen as forming a right-angle intersection with the vertical lines that are slanted away in depth from the observer, so that the lower ends of the slanted lines are seen as closer than the upper. When we consider the minimal form of the illusion, which has one slanted line intersecting a vertical line (Figure 3C), we find that only 12% of the responses imply any phenomenal depth. All of these, however, do imply the appropriate phenomenology with the slanted line seen tilted in space, as in "a religious cross seen from the side." When more inducing elements are present, as in Figure 3B, the expected depth relationships readily manifest themselves. For this configuration, 79% of the responses imply a depth relationship of the appropriate type, as in "an aerial seen from the side." Again, it seems as if the minimal format does not provide enough suggested depth to consistently elicit a phenomenal impression of depth. However, like the Ponzo, increasing the number of instances of the assumed depth cue dramatically increases the number of reports of perceived depth relationships.

The Orbison illusion (Figures 2E, 2F, 2G, and 2H). For both of the Orbison inducing configurations, consisting of a series of concentric circles (Figure 2G) and lines diverging to a starburst (Figure 2H), the phe-

nomenal depth that predicts the illusion requires that the center be seen as more distant. For the concentric-circle configuration, 32% of the responses imply some depth relationship. Of these, 97% imply the center as being seen to be more distant, as in "looking into a long tunnel." For the starburst configuration, however, only 6% of the responses imply any depth relationships, and all of these suggest that the center is seen as closer than the periphery, as in "a teepee viewed from the top."

The Sander parallelogram (Figures 2I and 2J). Of all of the configurations, the data for the Sander parallelogram are the most consistent with a constancy-scaling hypothesis. Here, we find 52% of the responses indicating phenomenal depth. In all of these, the configuration was seen as representing a rectangle viewed obliquely and receding into space from left to right as in "a football field viewed from the side." This is precisely the depth relationship that is required to predict this illusion.

The Delboeuf illusion (Figures 1E and 1F). In these configurations, there are no converging line elements or suggested texture gradients that would trigger depth relationships. Thus, it seems unlikely that many depth responses would be elicited by these configurations. In fact, for both the apparently smaller central circle (Figure 1E) and the apparently larger central circle (Figure 1F), approximately one-third of the responses imply relative depth (39% and 30% respectively). For both of these configurations most depth responses imply that the central circle is seen as being closer than the outer circle (90% and 83% respectively) as in "a push-button."

General discussion

Let us look at the total pattern of results produced by all of these configurations. The mean percentage of depth responses across the 13 illusion configurations is 41%. However, the pattern of depth responses is not consistent. The Sander parallelogram is the one illusion in which a large percentage of depth responses, consistent with misapplied constancy scaling theory, are obtained. For both halves of the Mueller-Lyer illusion, there is a strong bias toward reporting the center test shaft as being closer than the wing-tips. This apparent bias tends to reduce the probability that the appropriate depth relationships will be perceived in the apparently longer half of the illusion, where the shaft must be seen as more distant than the wing-tips for the theory to predict the illusion. However, despite this unidirectional bias in reported phenomenal depth, there is a large difference between the percentage of depth responses seen as near (or far) for the apparently longer as compared with the apparently shorter half of the illusion. This difference is in the predicted direction.

For the basic Ponzo configuration and for the minimal Zoellner configuration, there seems to be little perceived depth. However, the addition of extra inducing elements leads to a massive increase in the number of perspective depth responses. Thus, there may be implicit depth cues contained within the minimal configurations. However, unless these cues are made quite salient, they are not sufficiently strong to elicit phenomenal depth, although they may be strong enough to register as depth cues and thus trigger a constancy-scaling mechanism. It is interesting to note, in light of the extensive literature on constancy scaling in the Ponzo illusion and the almost total lack of literature on constancy scaling in the Zoellner illusion, that it is the normal form of the Ponzo that fails to elicit many depth responses while the normal form of the Zoellner elicits a very large percentage indeed.

For the Orbison configurations, one pattern produces the expected relationships while the other one does not. For the Poggendorff and horizontal-vertical illusions, the expected phenomenal depth does not appear And, as expected, appropriate differences in direction of depth reports do not appear in the Delboeuf illusion.

These results seem somewhat inconsistent with proposals that misapplied constancy scaling, based on implicit depth cues, can account for a large number of visual illusions (Gregory, 1966; Day, 1972). For the Sander parallelogram, the Mueller-Lyer illusion, the normal form of the Zoellner illusion, the Ehrenfels form of the Ponzo illusion and one of the Orbison configurations, the results are consistent with a misapplied constancy-scaling hypothesis. However, for the other configurations the results fail to support such a hypothesis. These data may possibly represent another instance of the difficult problem of *phenomenal* vs. *registered* depth, which has long puzzled researchers in the area of perceptual constancies. While alterations in the cues to distance seem to lead to predictable changes in apparent size, many investigators have failed to find the concomitant changes in phenomenal depth that would be expected with changes in the cues to distance (Wheatstone, 1852; Gruber, 1954; Hermans, 1954; Heinemann, Tulving, and Nachmias, 1959; Rock and McDermott, 1964; Gogel, Wist, and Harker, 1963; Rock and Kaufman, 1962). Thus, a stimulus configuration may fail to evoke apparent changes in depth while providing sufficient cues to evoke the appropriate constancy-scaling mechanisms.

Nonetheless, given the fact that the subjects in this experiment were instructed to view the configurations as pictorial stimuli, the absence of depth responses consistent with misapplied constancy-scaling theory for the basic form of the Ponzo, the minimal form of the Zoellner, the Pog-

gendorff, the horizontal-vertical, and the one variant of the Orbison illusion, are disturbing. It seems that the implicit depth cues required by the theory are quite inaccessible to conscious phenomenology since they are not available to observers attempting to interpret the stimuli as analogues of real-world configurations.

Notes

The authors would like to acknowledge the assistance of Fran Theodor in the coding and scoring of these results. This research was supported in part by grants from the National Research Council of Canada (A-9783, A-9958) and the National Science Foundation (74-18599). Requests for reprints should be sent to Lawrence M. Ward, Department of Psychology, University of British Columbia, Vancouver, B.C., V6T 1W5. Received for publication August 17, 1976; revision, November 3, 1976.

References

Coren, S., and Festinger, L. 1967. An alternative view of the "Gibson Normalization Effect." *Perception and Psychophysics* 2:621–626.

Coren, S., and Girgus, J. S. 1975. A size illusion based upon a minimal interposition cue. *Perception* 4:51–54.

Day, R. H. 1972. Visual spatial illusions: a general explanation. *Science* 175: 1335–1340.

Gilliam, B. 1971. A depth processing theory of the Poggendorff illusion. *Perception and Psychophysics* 10:211–216.

Girgus, J. S., and Coren, S. 1975. Depth cues and constancy scaling in the horizontal-vertical illusion: The bisection error. *Canadian Journal of Psychology* 29:59–60.

Gogel, W. C., Wist, E. P. R., and Harker, G. S. 1963. A test of the invariance of the ratio of perceived size to perceived distance. *American Journal of Psychology* 76:537–553.

Green, R. T., and Hoyle, E. M. 1963. The Poggendorff illusion as a constancy phenomenon. *Nature* 200:611–612.

Gregory, R. L. 1963. Distortion of visual space as inappropriate constancy scaling. *Nature* 199:678–680.

Gregory, R. L. 1966. *Eye and brain.* London: Weidenfeld and Nicolson.

Gregory, R. L. 1968. Visual illusions. *Scientific American* 219:66–76.

Gruber, H. E. 1954. The relation of perceived size to perceived distance. *American Journal of Psychology* 67:411–426.

Heinemann, E. G., Tulving, E., and Nachmias, J. 1959. The effect of ocular motor adjustments on apparent size. *American Journal of Psychology* 72: 32–45.

Hermans, T. G. 1964. The relationship of convergence and elevation changes to judgments of size. *Journal of Experimental Psychology* 48:204–208.

Hotopf, W. H. 1966. The size constancy theory of visual illusion. *British Journal of Psychology* 57:307–318.

Kilbride, P. L., and Leibowitz, H. W. 1975. Factors affecting the magnitude of the Ponzo illusion amongst the Baganda. *Perception and Psychophysics* 17: 543–548.

Leibowitz, H. W., Brislin, R., Perlmutter, L., and Hennessey, R. 1969. Ponzo perspective illusion as a manifestation of space perception. *Science* 166: 1174–1176.

Leibowitz, H. W., and Pick, H. A. 1972. Cross-cultural and educational aspects of the Ponzo perspective illusion. *Perception and Psychophysics* 12:430–432.

Pike, A. R., and Stacey, B. C. 1968. The perception of luminous Mueller-Lyer figures and its implications for the misapplied constancy theory. *Life Science* 7:355–362.

Rock, I., and Kaufman, L. 1962. The moon illusion II. *Science* 136:1031–1032.

Rock, I., and McDermott, W. P. 1964. The perception of visual angles. *Acta Psychologica* 22:119–134.

Thiery, A. 1896. Uber geometrisch-optische Tauschungen. *Philosophische Studien* 12:67–126.

Wheatstone, C. 1852. On some remarkable and hitherto unobserved phenomena of binocular vision: Part II. *Philosophical Magazine* Series IV, 504–523.

Worrall, N. 1974. A test of Gregory's theory of constancy scaling. *American Journal of Psychology* 87:505–510.

AMERICAN JOURNAL OF PSYCHOLOGY
December 1977, Vol. 90, No. 4, pp. 621–633

Spontaneous alternation in mice: A test of the mere-exposure hypothesis

G. J. Syme and Lesley A. Syme
University of Western Australia

Spontaneous alternation after confinement (habituation) to an arm of a T-maze was studied in three strains of albino mice: Simpson inbred, BALB/c, and Prince Henry random-bred. The confinements were from 5 min to 8 hr long, in a series of experiments with controls for both response bias and handling. As confinement increased, so did a tendency — related to sex, strain, and perhaps age — to prefer familiarity. After 8 hr of confinement, Prince Henry random-bred female mice showed a marked preference for familiarity. The findings are consistent with Zajonc's mere-exposure hypothesis and suggest further work on the relationship between site attachment and exploratory behavior.

At present, there appears to be some conflict between the various theories of exploration (e.g., Berlyne, 1967; Dember and Earl, 1957; Fiske and Maddi, 1961) and Zajonc's mere-exposure hypothesis (Zajonc, 1968, 1971). Generally, experimental data supporting the theories of exploration show the tendency of an organism to approach novel or complex stimuli, whereas the mere-exposure studies attempt to demonstrate a tendency to approach familiar stimuli.

Although the mere-exposure hypothesis has had some success in describing the responses of humans to a variety of stimuli (Heingartner and Hall, 1974; Saegert and Jellison, 1970; Zajonc, Reimer, and Hausser, 1973; Zajonc, Markus, and Wilson, 1974), there are some studies that seem to directly contradict it. One of the most obvious examples is found in the literature on spontaneous alternation, which has provided considerable support for the theories of exploration.

In one such study, Glanzer (1953) confined (habituated) rats to their chosen arm of a T-maze for 10 min before returning them to the startbox for a second free-choice trial. He found that rats confined to the chosen arm alternated significantly more often in the subsequent trial than rats confined for the same period to the startbox. These findings, since demonstrated in at least one other species (Hughes, 1967), led Glanzer to pro-

pose his stimulus-satiation hypothesis of spontaneous alternation (1958), although there are others (Dember and Fowler, 1958; O'Connell, 1965a).

Despite the studies of spontaneous alternation and other demonstrations of a preference for novel environs after confinement (e.g., Hughes, 1965, 1968), there is also considerable evidence of differing types of site attachment in animals (e.g., territorial and home-range behaviors). So far, it has been assumed by most authors that such spatial behavior is largely socially mediated (e.g., McBride, 1971). However, it has also been suggested that long-term exposure to a certain area can be at least a partial determinant of consistent spatial behavior (Pontius, 1967; Dimond, 1970, p. 150), thus providing some support for the mere-exposure hypothesis.

Considering the territorial and home-range theories, it is interesting that the studies of alternation after active or passive confinement have adopted relatively short periods of confinement (the longest, perhaps, being the 50-min confinement in the study by Egger, Livesey, and Dawson, 1973). It may be, therefore, that preferences for novelty rather than familiarity were observed in these studies because the confinements were not long enough to effect a preference for familiarity. That is, if the confinement is of sufficient length, Zajonc's hypothesis may be tenable in this test situation.

The present experiments were designed to test that hypothesis. Mice were chosen as subjects because these animals have been shown to alternate spontaneously (Petchkovsky and Kirkby, 1970), to explore after confinement (Hughes, 1969), and experimentally, to exhibit territoriality (Mackintosh, 1970, 1973).

EXPERIMENT I

This experiment was conducted to examine the effects on alternation of confinement to the chosen arm of a T-maze for 30 min as compared to confinement in the startbox for the same period.

METHOD

Subjects

The subjects were 16 naive male mice (Simpson albino inbred strain) about 150 days old and housed in pairs in cages measuring 27 by 15 by 11 cm for a month before the experiment began. Food and water were freely available.

Apparatus

The apparatus was a wooden T-maze with independently hinged Plexiglas lids on the start alley and goal arms. The start alley was 60 by 9 by 7.5 cm and was painted dark gray. Each goal arm measured 40 by 9 by 7.5 cm. For the

initial observations one goal arm was dark gray and the other white, but in the later observations both arms were dark gray. Similarly colored sliding metal barriers at the point of entry to each goal arm were used to prevent retracing or escape during confinement.

Procedure

A repeated-measures design was employed (see Hughes, 1967) in which each mouse was run for 16 pairs of trials. Each day every mouse was given two trials in the maze. After its first choice of a goal arm, it was confined to that arm or to the startbox for 30 min before it was released again from the startbox for its second trial. The maze was not cleaned between trials, but it was wiped with a damp cloth between subjects. For half the 16 days the mice were confined to their chosen arm for 30 min after the first trial, and for the other half they were confined to the startbox (see Glanzer, 1953). Half the mice were confined first to their chosen arm and then to the startbox, and the others in the opposite order.

A choice between white and dark gray goal arms was provided in the first 8 days in order to approximate Glanzer's experimental conditions. However, because of the results obtained, both goal arms were dark gray for the final 8 days.

A trial was terminated if the mouse had not entered a goal arm after 10 min (Petchkovsky and Kirkby, 1970).

RESULTS

For days 1–8 the white goal arm was chosen over the dark gray one in only 8 of the 128 choices. Because of this marked preference, two dark gray goal arms were used thereafter.

For days 9–16, the numbers of times each mouse alternated after confinement to its chosen arm or to the startbox were compared using a t test for related samples. No significant difference was found between the means [for the chosen arm, 1.81; for the startbox, 1.93, $t = .50, p > .05$]. An estimate of the expected level of alternation (Manning, 1973) on the first trial on each of these days was obtained using

$$pA = (k/N)(1 - k/N)(N/N - 1).$$

Deviation scores from this expected value were obtained for each mouse for trials after confinement to its chosen arm and for trials after confinement to the startbox. A one-sample t test was then used to test the null hypothesis that the mean difference beween observed and expected alternations was zero under both conditions. After confinement to the chosen arm, the resultant t value was not significant [$t(15) = 1.88, p > .05$]. A nonsignificant result was also obtained after confinement to the startbox [$t(15) = .47, p > .05$]. Thus, there was no significant tendency for the mice to alternate under either condition.

Unlike previous findings with rats (Glanzer, 1953), then, and in comparison with control values, there was no increase in the alternation scores of mice that had been confined to their chosen goal arm. In fact, the mice showed no tendency to alternate in either situation. While this is not surprising after confinement to the startbox, in that Walker (1956) found decreased alternation in rats as the intertrial interval increased, alternation behavior has been shown to be retained for up to 18 hr (O'Connell, 1965b).

It is interesting that alternation also failed to occur after the mice were confined to their chosen goal arm. This behavior might represent the middle stage of a growing preference for the familiar side of the apparatus, or it could be an artifact of the large number of trials, during which the mice may have become familiar with both goal arms. Experiment II was conducted to distinguish between these two possibilities.

EXPERIMENT II

In this experiment, alternation in mice was compared over four lengths of confinement to their chosen arm, in order to demonstrate the effects of increasing familiarity on subsequent alternation. On this occasion, the mice were not confined to the startbox as a control. This was to prevent excessive familiarization with the apparatus as a whole.

METHOD

Subjects and apparatus

The subjects were 32 mice, 14 females and 18 males about 120 days old and of the Simpson albino inbred strain (see Hughes, 1969). They were housed either in single-sex pairs in cages measuring 27 by 15 by 11 cm or in three single sex groups of six in cages measuring 24 by 18 by 13 cm. Food and water were freely available. The apparatus was the same as that employed for days 9–1[illegible] in Experiment I. For purposes of identification, each mouse had its tail colored with a commercial dye.

Procedure

Each mouse was run for two trials daily, but on this occasion there were four lengths of confinement to the chosen arm: 0 min (no confinement, the mouse replaced in the startbox as soon as possible after its choice of a goal arm) 5 min, 30 min, and 60 min. Each mouse was tested once at each length of confinement, giving four measures of alternation in all. To control for order effects a randomized design was adopted, with the limitation that each length of confinement appear equally often in each position.

RESULTS

The percentage of alternations after each length of confinement is shown in Figure 1. All four sets of frequencies were subjected to a Cochran Q test before individual comparisons were made. The resultant value [$Q(3) = 15.87, p < .01$] indicated that the four lengths of confinement had differential effects on alternation. A series of individual McNemar tests for significance of changes were then calculated. Frequency of alternation after 60 min of confinement proved to be significantly less than after 30 min [$\chi^2(1) = 8, p < .01$] or 5 min of confinement [$\chi^2(1) = 11.84, p < .001$]. The only other significant individual comparison showed that alternation after 30 min of confinement was significantly less than after 5 min of confinement [$\chi^2(1) = 4, p < .05$].

The first daily trials of each subject were examined for response bias (Douglas, 1966). None were found: of the 128 first daily trials 54% were right and 46% left. Using Douglas and Isaacson's (1965) formula for the expected frequency of alternation, the resultant value was .50.

A chi-square test calculated within each length of confinement revealed that there was a significant tendency to alternate after 5 min of confinement [$\chi^2(1) = 8, p < .01$], whereas after 60 min of confinement, there was a significant tendency to perseverate [$\chi^2(1) = 6.1, p < .02$]. However, as illustrated in Figure 2, an analysis of the separate results for males and

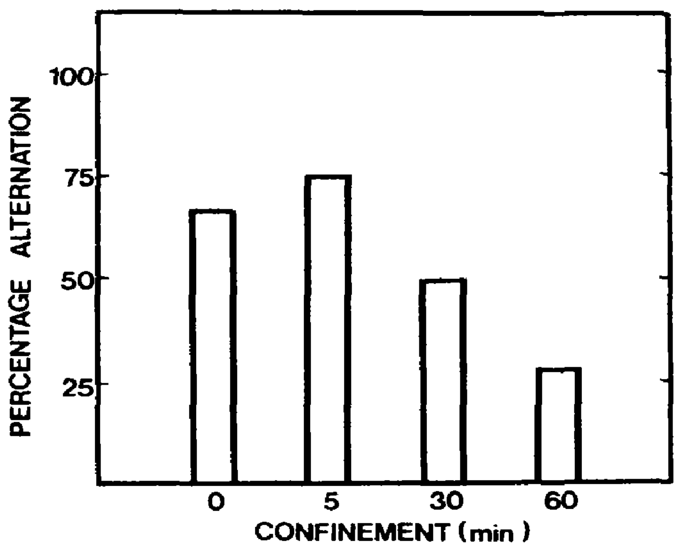


Figure 1. The percentage of alternation after each length of confinement for both sexes ($N = 32$); Experiment II

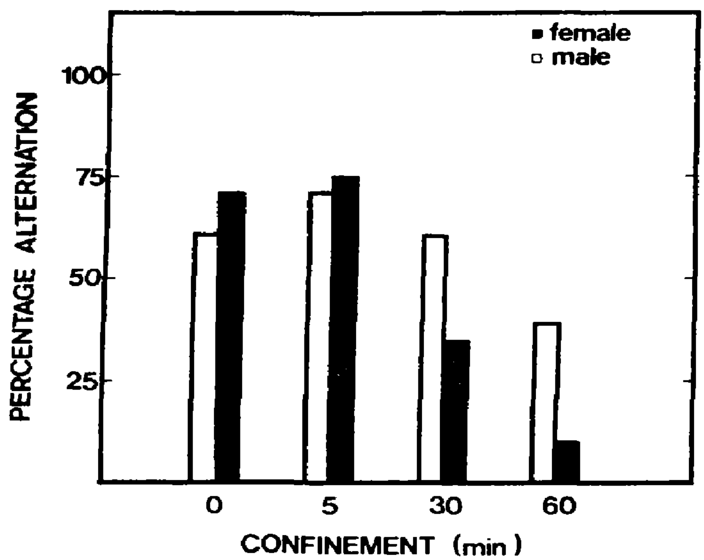


Figure 2. The percentage of alternation after each length of confinement for males ($N = 18$) and females ($N = 14$); Experiment II

females after 60 min of confinement showed a significant preference for familiarity by the females [$\chi^2(1) = 7.14, p < .01$], but not by the males [$\chi^2(1) = .44, p > .05$].

Finally, the 65% degree of alternation with no prior confinement was slightly lower than the 74.6% Petchkovsky and Kirkby (1970) found using a repeated-measures design (74.6%).

Thus, the results support Zajonc's mere-exposure hypothesis, especially for the females. A slight, though nonsignificant, increase in alternation was observed after 5 min of confinement as compared to no confinement This is comparable to earlier findings with rats (Glanzer, 1953). Further confinement, however, considerably reduced alternation; there was an exactly chance rate of alternation after 30 min of confinement and a significant tendency to perseverate after 60 min of confinement. For mice in a T-maze, then, prolonged exposure to a goal arm appears to promote an affiliative response to that arm. But there may be a problem with this interpretation. Walker (1956) has shown that if very long intertrial intervals are used (i.e., +7 hr), there is a slight tendency for rats to perseverate. It is possible, therefore, that it was the length of the intertrial interval that caused the perseveration by the mice in the present study, regardless of their area of confinement. Experiment III was conducted to investigate this possibility.

EXPERIMENT III

For this experiment, the mice were again confined for 60 min between trials. On this occasion, however, they were confined to the startbox.

METHOD

Subjects, apparatus, and procedure

The subjects and apparatus were as for Experiment II. The mice were rested for four days and then run for two trials in the T-maze. After recording each animal's choice of a goal arm on the first trial, the animals were confined for 60 min to the startbox before they were released for the second trial.

RESULTS

Under these conditions, 17 of the 32 mice alternated. This rate of alternation is comparable to that obtained after 30 min of confinement to the startbox in Experiment I. The rate of alternation after 60 min of confinement to the startbox here and that after 60 min of confinement to the chosen arm in Experiment II were compared using separate-sex McNemar tests for significance of changes. The resultant value was statistically significant for the females [$\chi^2(1) = 4, p < .05$]. However, only four males changed their response from the goal box habituated condition, and though all changed in the direction of alternation, a nonsignificant result was obtained.

The perseveration observed for the females in Experiment II was thus a result of confinement to a specific area rather than a general increase in intertrial interval.

EXPERIMENT IV

The results so far, especially for the female mice, provide some support for the mere-exposure hypothesis. However, one fault in the design of these studies could be the use of different handling procedures with and without confinement in Experiment II. Whereas the nonconfined mice were handled almost immediately after they entered their chosen goal arm, the confined mice were not handled until that period of confinement had elapsed. Thus, the nonconfined animals may have associated the chosen arm with the aversive stimulation of human handling and subsequently avoided repetition of their first choices more often than the confined mice. Experiment IV was designed to control for this factor by handling all of the mice as soon as they entered a goal arm.

An independent-measures design was also adopted to completely control for the possible contaminating effect of repeated exposure to the test environment. Finally, two strains of mice were used, since genetic variables can influence alternation in this species (Henderson, 1970).

METHOD

Subjects and apparatus

The 96 subjects were all female albino mice 70 days old, 48 of them of the BALB/c inbred strain and 48 of them of the Prince Henry random-bred strain. All animals were housed in groups of six with food and water freely available. The apparatus was the same as that used in previous experiments.

Procedure

The mice were randomly assigned (16 per group) to three lengths of confinement: 0 hr, 1 hr, or 2 hr. On entering a goal arm, each mouse, regardless of its condition for length of confinement, was handled for 5 sec by the experimenter.

RESULTS

The number of mice alternating after each length of confinement is shown in Table 1. An independent chi-square analysis revealed a significant tendency for the BALB/c mice to alternate less as confinement increased [$\chi^2(2) = 6.9, p < .05$]. This effect was not observed for the Prince Henry strain [$\chi^2(2) = .09, p > .05$].

Decreased alternation with increased confinement was therefore replicated in the BALB/c strain when the confounding variable of differential handling was controlled. But even here, the level of alternation was only slightly below 50% after a 2-hr confinement. No significant response bias was observed for the BALB/c mice; therefore, no significant preference for novelty or familiarity occurred. The mere-exposure hypothesis was not unequivocally supported.

The similarity between the present findings and those on the effects of

Table 1. The number of mice choosing the novel arm ($N = 16$ per cell); Experiment IV

	Confinement (hr)		
Strain	0	1	2
BALB/c	14	9	7
Prince Henry	9	10	10

increasing intertrial interval on alternation without confinement (Walker, 1956; O'Connell, 1965b, 1971) is striking. Those studies report a U-shaped relationship between the level of alternation and the length of the intertrial interval. After short delays alternation decreases; a longer delay between trials produces a gradual increase in this behavior. It appears that, as in Experiment I, the early drop in alternation can be observed even when the subject is habituated to a goal arm.

One explanation of this similarity could be that a tendency to repeat the most recent response increases in the short term, regardless of the novelty or familiarity of the goal arms. A final experiment was conducted to differentiate between the effects of habituation on response bias and the preference for novelty or familiarity.

EXPERIMENT V

Earlier studies (e.g., Montgomery, 1952) have achieved a distinction between response bias and the preference for novelty or familiarity by using a +-maze in which the animals were run from different ends on separate trials. In order to approach the novel arm, the animal had to turn the same way as for the previous trial, so that novelty and response bias were clearly defined. In the present situation, however, since either a preference for familiarity or novelty could be hypothesized, half of the mice were habituated to the goal arm they chose to enter on their first trial and the other half to the goal arm they did not choose to enter on their first trial.

A further departure from the conventional procedure involved allowing the mice access to the stem of the start alley of the maze during habituation to avoid having the mice 'forget' how they responded. The early drop in alternation has been described in terms of the Kamin effect (O'Connell, 1971).

METHOD

Subjects and apparatus

The subjects were 48 female BALB/c mice about 100 days old when tested and 48 Prince Henry female mice about 120 days old when tested. Housing conditions and apparatus were the same as for the previous study.

Procedure

Three lengths of 'confinement' (more precisely, habituation) were used: 2 hr, 5 hr, and 8 hr. Sixteen animals of each strain were then assigned to each condition. The mice in all conditions were placed in the startbox. They were then

released and, once a goal arm was entered, returned in less than 30 sec to the startbox. When the mice were put in the startbox for the second time, the door leading to one of the goal arms was closed. For half the mice in each condition, the goal arm the mouse entered on the first trial was blocked, while for the rest, it was the goal arm the mouse did not enter. Thus, half the mice were habituated to their chosen arm and half to their nonchosen arm. As soon as the door was shut, the mice were released from the startbox. After the allotted confinement period had elapsed, the observer waited until the mouse was in the startbox arm of the maze and returned it to the startbox; if the mouse was already in the startbox, it was kept there but was also picked up to control for handling effects. The doors of both goal arms were raised and the goal arm then entered by each mouse was recorded.

RESULTS

The number of mice in each condition to choose the novel arm can be seen in Table 2. The mice of the BALB/c strain showed a uniform slight preference (34 of the 48 mice) for novelty [$\chi^2(1) = 8.33, p < .01$]. Those of the Prince Henry strain, however, preferred the familiar arm after the two longer confinements [$\chi^2(2) = 7.37, p < .05$]: only 9 of 32 chose the more novel arm after 5- and 8-hr confinements [$\chi^2(1) = 6.13, p < .02$].

The results can also be analyzed in terms of the initial response tendency of the animals. Only one cell, however, showed any tendency to deviate from chance values. For BALB/c mice after 2 hr of confinement, only 4 animals alternated responses [$\chi^2(1) = 4.0, p < .05$]. This finding should be interpreted cautiously, though, since there were five other nonsignificant cells in the analysis.

DISCUSSION

In Experiment V the mice of one strain supported the mere-exposure hypothesis, while those of the other strain demonstrated the common preference for novelty. The results also indicate that the habituation effects observed for the BALB/c mice in Experiment IV could well be attrib-

Table 2. The number of mice choosing the more novel arm ($N = 16$ per cell); Experiment V

Strain	'Confinement' (hr)		
	2	5	8
BALB/c	12	11	11
Prince Henry	11	4	5

uted to increased response perseveration after 2 hr of confinement. Further work is required to answer this question.

In all experiments, then, some support was obtained for the mere-exposure hypothesis. Choice for familiarity seems to depend on habituation, strain, sex and perhaps even age. A systematic investigation of these variables will illustrate the relative importance of each.

Thus, in the absence of any social paradigm, site attachment can occur. Whether this attachment can wholly explain territorial or home-range behavior is another matter. Female mice, which seemed to be more responsive than males to habituation in the present experiments, are not generally regarded as being territorial except, perhaps, during reproduction (Noirot, 1969). Social mechanisms appear to play a vital role in many species for the regulation of spatial behavior (see Fisler, 1969 for review). Our experiments suggest that nonsocial site attachment may also play an important part in determining this behavior.

An even more perplexing theoretical problem arises when we consider the relationship between preference for familiar and novel environments or between the mere-exposure effect and exploration. Hughes (1969) found exploratory behavior in mice when he gave his animals a choice between novel and familiar environs after confining them to the familiar side for 23 hr. This would seem to contrast with the present findings. In further contrast, Furchtgott and Cureton (1964) confined mice to a cage for as little as 48 hr before lifting the lid to allow them to emerge. The latency to escape in this situation was used as a measure of emotionality and has since been labeled 'home-cage emergence' (Archer, 1973). But when does a cage become regarded by its occupant as a home? And when does exploration give way to site attachment, or vice versa?

Although the reasons for the preference for familiarity cannot be resolved in this discussion, it is evident that Zajonc's mere-exposure hypothesis can be supported in spatial terms. Since this is the case, there is urgent need for additional experimental evidence to establish the relationship between site attachment and exploratory behavior.

Notes

Address requests for offprints to Dr. G. J. Syme, Division of Land Resources Management, CSIRO, Wembley, W.A. 6014, Australia. Received for publication October, 1974; revision, January, 1976.

References

Archer, J. 1973. Tests for emotionality in mice and rats: A review. *Animal Behaviour* 21:205–235.

Berlyne, D. E. 1967. Arousal and reinforcement. In *Nebraska symposium on motivation, 1967,* ed. D. Levine. Lincoln: University of Nebraska Press.

Dember, W. N., and Earl, R. W. 1957. Analysis of exploratory, manipulatory, and curiosity behaviors. *Psychological Review* 64:91–96.

Dember, W. N., and Fowler, H. F. 1958. Spontaneous alternation behavior *Psychological Bulletin* 55:412–428.

Dimond, S. J. 1970. *The social behavior of animals.* London: Batsford.

Douglas, R. J. 1966. Cues for spontaneous alternation. *Journal of Comparative and Physiological Psychology* 62:171–183.

Douglas, R. J., and Isaacson, R. L. 1965. Homogeneity of single-trial response tendencies and spontaneous alternation in the T-maze. *Psychological Reports* 16:87–92.

Egger, G. J., Livesey, P. J., and Dawson, R. G. 1973. Ontogenetic aspects of central cholinergic involvement in spontaneous alternation behavior. *Developmental Psychobiology* 6:289–299.

Fiske, D. W., and Maddi, S. R., eds. 1961. *Functions of varied experience* Homewood, Ill.: Dorsey.

Fisler, G. F. 1969. Mammalian organisation systems. *Los Angeles County Museum Contributions to Science* 167:1–32.

Furchtgott, E., and Cureton, E. 1964. Factor analysis of emotionality and conditioning in mice. *Psychological Reports* 15:787–794.

Glanzer, M. 1953. The role of stimulus satiation in spontaneous alternation. *Journal of Experimental Psychology* 45:387–393.

Glanzer, M. 1958. Curiosity, exploratory drive, and stimulus satiation. *Psychological Bulletin* 55:302–315.

Heingartner, A., and Hall, J. V. 1974. Affective consequences in adults and children of repeated exposure to auditory stimuli. *Journal of Personality and Social Psychology* 29:719–723.

Henderson, N. D. 1970. A genetic analysis of spontaneous alternation in mice *Behavior Genetics* 1:125–132.

Hughes, R. N. 1965. Food deprivation and locomotor exploration in the white rat. *Animal Behaviour* 13:30–32.

Hughes, R. N. 1967. Effects of intertrial confinement and brightness change on ferret alternation behavior. *Psychological Reports* 20:679–681.

Hughes, R. N. 1968. Behavior of male and female rats with a free choice of two environments differing in novelty. *Animal Behaviour* 16:92–96.

Hughes, R. N. 1969. Exploration in three laboratory rodents. *Perceptual and Motor Skills* 28:90.

Mackintosh, J. H. 1970. Territory formation by laboratory mice. *Animal Behaviour* 18:177–183.

Mackintosh, J. H. 1973. Factors affecting the recognition of territory boundaries by mice (*Mus musculus*). *Animal Behaviour* 21:464–470.

Manning, S. K. 1973. A design comparing free and forced-choice alternation in the rat at two interchoice intervals. *Behavior Research Methods and Instrumentation* 5:263–265.

McBride, G. 1971. Theories of animal spacing: the role of flight, fight, and social distance. In *Behavior and environment,* ed. A. H. Esser. New York: Plenum Publishing.

Montgomery, K. C. 1952. A test of two explanations of spontaneous alternation. *Journal of Comparative and Physiological Psychology* 45:287–293.

Noirot, E. 1969. Interactions between reproductive and territorial behavior in female mice. *International Mental Health Research News* 11:10–11.

O'Connell, R. H. 1965a. Trials with tedium and titillation. *Psychological Bulletin* 63:170–179.

O'Connell, R. H. 1965b. Spontaneous alternation following long intertrial intervals. *Psychological Reports* 17:131–134.

O'Connell, R. H. 1971. Delayed alternation in rats with position preferences. *Psychonomic Science* 22:137–139.

Petchkovsky, L., and Kirkby, R. J. 1970. Individual differences, emotionality, and spontaneous alternation in mice. *Australian Journal of Psychology* 22:75–78.

Pontius, A. A. 1967. A neuropsychiatric hypothesis about territorial behavior. *Perceptual and Motor Skills* 24:1232–1234.

Saegert, S. C., and Jellison, J. M. 1970. Effects of initial level of response competition and frequency of exposure on liking and exploratory behavior. *Journal of Personality and Social Psychology* 16:553–558.

Walker, E. L. 1956. The duration and course of the reaction decrement and the influence of reward. *Journal of Comparative and Physiological Psychology* 49:167–176.

Zajonc, R. B. 1968. The attitudinal effects of mere exposure. *Journal of Personality and Social Psychology Monograph Supplement:* 9(2, part 2): 1–27.

Zajonc, R. B. 1971. Attraction, affiliation and attachment. In *Man and beast: Comparative social behavior,* ed. J. F. Eisenberg and W. S. Dillon. Washington, D.C.: Smithsonian Institute.

Zajonc, R. B., Reimer, D. J., and Hausser, D. 1973. Imprinting and the development of object preference in chicks by mere repeated exposure. *Journal of Comparative and Physiological Psychology* 83:434–440.

Zajonc, R. B., Markus, H., and Wilson, W. R. 1974. Exposure, object preference, and distress in the domestic chicken. *Journal of Comparative and Physiological Psychology* 38:667–678.

continued from page 608

Rosenbaum, Max, and Snadowsky, Alvin. *The Intensive Group Experience.* New York: Free Press, 1976. Pp. 210. $13.95.

Rosenthal, Robert. *Experimenter Effects in Behavioral Research.* New York: Halsted, 1976. Pp. 500. $15.95.

Rosenzweig, Mark R. *Biologie de la Memoire.* Paris: Presses Universitaire de France, 1976. Pp. 162. Paperback.

Rossman, Parker. *Sexual Experience between Men and Boys.* New York: Association Press, 1976. Pp. 247. $10.95.

Rothenberg, Albert, and Hausman, Carl R., eds. *The Creativity Question.* Durham, N.C.: Duke University Press, 1976. Pp. 366. $14.95.

Rubin, Lillian Breslow. *Worlds of Pain: Life in the Working-Class Family.* New York: Basic Books, 1976. Pp. 368. $11.95.

Rubin, Theodore Isaac. *Love Me, Love My Fool.* New York: David McKay, 1976. Pp. 94. $5.95.

Rue, James J., and Shanahan, Louise. *A Catechism for Divorced Catholics.* St. Meinrad, Indiana: Abbey Press, 1976. Pp. 66. $1.95.

Ruitenbeek, Hendrik M. *Psychotherapy: What It's All About.* New York: Avon, 1976. Pp. 160. Paperback. $1.95.

Rummel, R. J. *Understanding Conflict and War.* Vol. 2. *The Conflict Helix* New York: Halsted, 1976. Pp. 400. $17.50.

Russell, Diana E. H., and Van de Ven, Nicole. *Crimes against Women.* Millbrae, Calif.: Les Femmes, 1976. Pp. 294.

Russell, Harold E., and Beigel, Allan. *Understanding Human Behavior for Effective Police Work.* New York: Basic Books, 1976. Pp. 303. $13.95.

Russell, O. Ruth. *Freedom to Die.* New York: Dell, 1976. Pp. 334. Paperback. $1.95.

Safer, Daniel J., and Allen, Richard P. *Hyperactive Children: Diagnosis and Management.* Baltimore: University Park Press, 1976. Pp. 239. Paperback $8.50.

Sager, Clifford J. *Marriage Contracts and Couple Therapy.* New York: Brunner-Mazel, 1976. Pp. 335. $15.00.

Salzinger, Kurt, ed. *Psychology in Progress: An Interim Report.* New York. Annals New York Academy Sciences, 1976. Pp. 270.

Sampson, Edward E. *Social Psychology and Contemporary Society.* 2d ed New York: Wiley, 1976. Pp. 567. $12.95.

Sandstrom, C. I. *The Psychology of Childhood and Adolescence.* Baltimore: Penguin Books, 1976. Pp. 259. Paperback. $3.25.

Sarason, Irwin G. *Abnormal Psychology: The Problem of Maladaptive Behavior.* 2d ed. Englewood Cliffs, N.J.: Prentice-Hall, 1976. Pp. 687. $14.95.

Sarason, Irwin G., and Spielberger, Charles D. *Stress and Anxiety.* Vol. III New York: Wiley, 1976. Pp. 365. $18.95.

Sass, Gregory. *If I Tell You, Will I Feel Less Scared?* Toronto: McDonald House, 1976. Pp. 91. Paperback. $5.95.

Satir, Virginia. *Making Contact.* Millbrae, Calif.: Celestial Arts, 1976. Paperback. $3.95.

continued on page 688

AMERICAN JOURNAL OF PSYCHOLOGY
December 1977, Vol. 90, No. 4, pp. 635–644

Environmental and behavioral cues in the perception of social encounters: An exploratory study

Joseph P. Forgas and L. B. Brown
University of Oxford, England, and University of New South Wales, Australia

Slides showing realistic dyadic encounters in naturalistic settings, using different combinations of behavioral and environmental cues, were the stimuli. Subjects ($N = 60$) rated their perception of these encounters on six scales, and responses were analyzed by multivariate analysis of variance and a multiple-regression technique. Both behavioral and environmental cues had a significant effect on the ratings of the encounters; the interaction of these cues was also highly significant, with the effects of the behavioral cues attenuated or accentuated by the setting; and there were important scale-specific differences in the effects of the two types of cues. The relative importance of environmental cues in social perception is discussed.

The complex relationship between social behavior and the physical environment or *setting* in which that behavior takes place has been the object of increasing interest in recent years. The concept of such settings was first introduced by Barker (1968; Barker and Wright, 1955), who suggested that particular environments are associated with typical, recurring patterns of social behavior. From the nature of its setting, much ongoing behavior can be predicted without any further assumptions about the intrapsychic aims, motivations, or personalities of the social actors (Stern, Stein, and Bloom, 1956).

The symbolic value of environments in social behavior was most explicitly recognized by researchers working within the symbolic-interactionist paradigm. Bennett and Bennett (1970, p. 190) write, for example, that "all social interaction is affected by the physical container within which it occurs. The various elements of the container establish a world of meaning through the arrangement of nonverbal symbolism. For this reason, the common practice in the social sciences of focussing on behaviour without reference to the physical setting would seem to ignore an important dimension of the total picture of interaction." The intricate relationship be-

tween environment and social behavior in a medical setting was elegantly analyzed by Ball (1970), and 'regions' play an important part in many of Goffman's (1959, 1963) insightful analyses of public and private behavior. In this sense, settings are not mere physical environments; rather, they are well understood symbolic codes representing social norms and expectations about appropriate and inappropriate behaviors.

A slightly different conceptualization is offered by environmental psychologists (Craik, 1973; Goodey, 1971; Proshansky, Ittelson, and Rivlin, 1970), who are more concerned with the physical characteristics, 'atmosphere,' aesthetic and amenity values of different places. Empirical studies in this tradition typically measure subjective reactions to different environments, such as the perceived climates of large institutions (Pace and Stein, 1968). In a similar vein, Moos (1973, 1974) and his co-workers have studied the connection between setting and social behavior in small-scale environments, such as psychiatric wards.

In social psychology, only the most cursory attempts have been made to deal with the environmental dimension in interactive behavior and social perception. The present study is concerned mainly with the role physical environments play in defining social interactions for observers. That settings have such a definitional function in social perception follows not only from Barker's, Moos's, and others' understanding of the concept but also from accumulating evidence suggesting the culturally coded and stereotypical nature of many social encounters (Forgas, 1976; Triandis, 1972). Such recurring stereotypical interactions with implicitly understood behavioral expectations and consensual boundaries in time and space can be seen as constituting 'social episodes.' Social episodes are elements of the subjective culture (Triandis, 1972) and are consensually defined by the setting, the relationship and history of the interactors, and other cues (Altman and Lett, 1970; Barker, 1968; Goffman, 1959, 1963). The main objective of the present study was to demonstrate that different environments have a potent and nonobvious effect on the way different social encounters are perceived by observers.

To evaluate the function of settings in the perception of social episodes, two specific questions were examined. First, do variations in the environmental setting itself significantly affect the perceptions of a social interaction? And second, what are the respective weights of cues from the setting and from the nonverbal behavior of the interactors in the observer's perception of a social episode, and how does he use these cues in congruent and incongruent combinations of setting and observed behavior?

METHOD

Subjects

In the pilot study (prejudgment), subjects were 12 undergraduate students, male and female. In the main study, a second sample ($N = 60$) of both undergraduate students and students from a school of occupational therapy was used. As a preliminary analysis revealed no differences due to sex or sample, data for all subjects were pooled for subsequent analysis.

Stimulus materials

A series of photographs was taken of four interacting dyads, who were verbally instructed to enact a series of different levels of nonverbal interaction intimacy. In all conditions, the couples were standing at about 12 ft from the camera; nonverbal intimacy was varied in terms of cues such as distance, head and body orientation and inclination, smile, gaze, gestures, and open or closed position of the extremities. From the numerous photographs so prepared, for each dyad four distinct photos, representing the whole range from intimate to nonintimate nonverbal communication (NVC) were selected by a small panel and prerated by a pilot sample (see below).

In addition, a series of photographs of different sorts of environmental *settings* was taken, a series which included inside and open-air settings and public and private places as well. Altogether eighteen such photos were prepared, including settings such as railway stations, street scenes, libraries, living rooms, theater lobbies, dining rooms, market scenes, and so on. Each of these settings was also prerated in the prejudgment stage.

Scales

Instead of the usual single bipolar scale (Mehrabian, 1968), a set of six seven-point scales was used. These scales (warm/cold; simple/complex; passive/active; intimate/nonintimate; pleasant/unpleasant; and tense/relaxed) were selected after a careful review of the literature (Warr and Knapper, 1968; Goodey, 1975; Moos, 1974) for dimensions found to be generally relevant to the perception of both interactions and behavior.

Prejudgment

The four photographs for each of the four dyads and the eighteen photographs of behavior settings were prerated by the pilot sample of 12 subjects on the set of six seven-point bipolar scales. For each of the four dyads, two photographs, one of an intimate and the other of a nonintimate communication, were selected; these two, according to the ratings made in the pilot study, were comparable not only in level of intimacy but also in their rating profiles on all six scales. Similarly, two of the eighteen settings were selected so that the rating profiles maximally matched the mean rating profiles of the intimate and nonintimate interacting couples. These two settings were a section of one of the lobbies of the Covent Garden Opera House (warm, intimate; setting 1) and a section of an open-air street market (cold, nonintimate; setting 2). The matching of the settings and communications was necessary so that their respective contributions to judgments of combined scenes could be evaluated. The mean rating profiles of the dyads and the settings are presented in Figure 1, which shows an extremely good match between these two sets of cues.

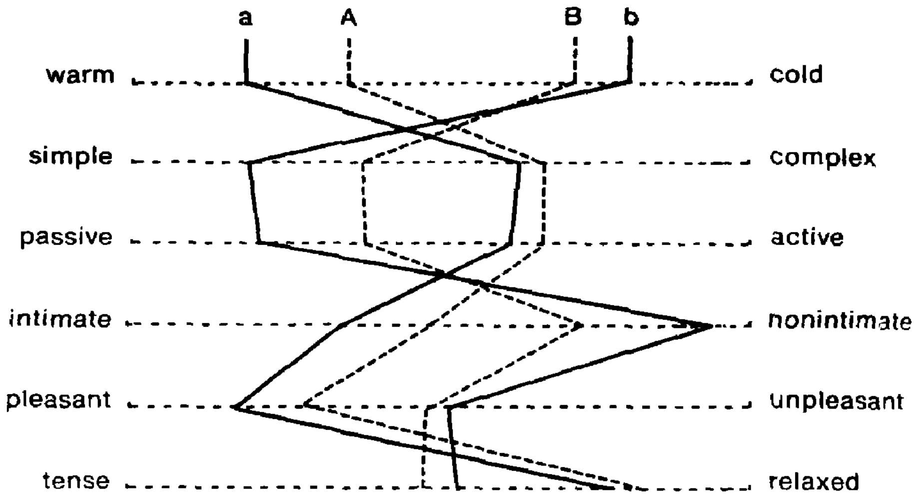


Figure 1. Mean rating profiles of behavioral and environmental cues: a, intimate communication; b, nonintimate communication; A, intimate setting; B, nonintimate setting (the standard deviations associated with these mean ratings ranged from .8 to 1.40)

Finally, intimate and nonintimate communications were superimposed on intimate and nonintimate settings for all couples and in all combinations, and the *combined scenes* ($N = 16$) were prepared as slides. It was possible to combine settings and interacting couples in such a way that a completely plausible and realistic picture of a real-life interaction in a natural setting was created.

Procedure and design

A complete within-subjects multivariate design was used, evaluating the effects of two independent factors (nonverbal behavioral communication and environmental setting) on a set of six dependent variables (scales). Since a preliminary analysis revealed no significant differences between the dyads, they were regarded as replications in the design.

Subjects were run in small groups of 10–12 persons. The order of presentation of the slides and of the scales was randomly varied across groups, in order to avoid series and response-set effects.

On the subjects' arrival, a questionnaire booklet with written instructions was handed out to them, the booklet describing the task as concerned with the "judgement of social situations." The slides were then presented in a semidarkened room, allowing enough time for the subjects to fill in the questionnaires after each slide.

RESULTS

A multivariate analysis of variance (MANOVA) was carried out, testing the overall effects of environmental setting (two levels) and nonverbal behavioral communication (two levels) on six bipolar scales (dependent

variables). This analysis showed that both the environmental cues [$F(6, 235) = 3.96, p < .001$] and the behavioral cues [$F(6, 235) = 28.8, p < .001$] had a significant effect on the observers' ratings of the combined scenes. The interaction of these two factors was also highly significant [$F(6, 235) = 13.11, p < .001$].

A more detailed, scale-by-scale analysis of the two environmental settings and two behavioral communications (univariate analyses of variance) indicated some interesting scale-specific differences: the results of these analyses are summarized below, while the mean cell ratings are shown in Figure 2. Intercorrelations between the scales indicated the presence of three clusters: the 'warm,' 'intimate,' and 'pleasant' scales (evaluative); the 'simple' and 'passive' scales; and the 'tense' scale.

Environmental *setting* significantly affected ratings on two scales: the episode was judged to be more warm [$F(1, 240) = 7.29, p < .001$] and relaxed [$F(1, 240) = 6.19, p < .001$] when it took place in the more intimate setting.

Behavioral *communication* significantly affected four scales, indicating

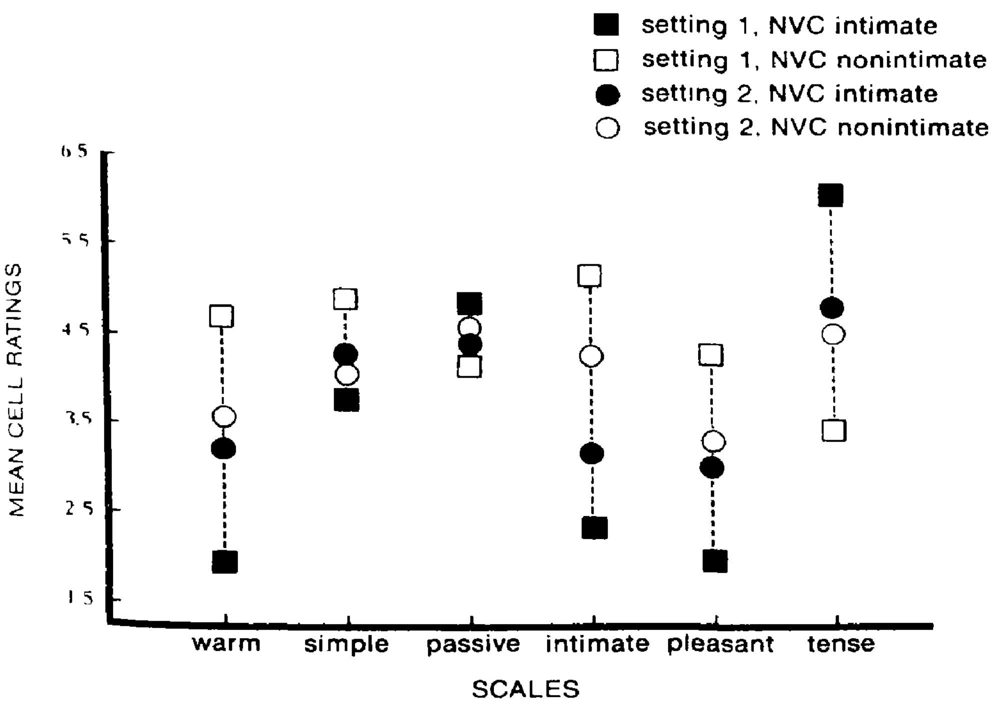


Figure 2. A graphic analysis of the interaction of behavioral and environmental cues in the perception of social episodes: Mean cell ratings on the six rating scales

that the more intimate communication resulted in the encounters being judged more warm [$F(1, 240) = 6.13$, $p < .001$], intimate [$F(1, 240) = 13.52$, $p < .001$], pleasant [$F(1, 240) = 15.04$, $p < .001$], and relaxed [$F(1, 240) = 16.69$, $p < .001$].

There were significant interactions of setting and communication on the same four scales however [$p < .001$], suggesting that the effects of the behavioral cues on the perception of the encounter could have been strongly influenced by the setting. Indeed, a graphic analysis of cell means (Figure 2) clearly shows that differences between behavioral and environmental cues were salient only in setting 1, but not in setting 2.

These results, on the whole, indicate that even though settings and communications were matched on all six scales, their effect on the observers' ratings of the combined scenes was interactive rather than additive or multiplicative. In order to quantify further the respective contributions of behavioral and environmental cues to the perception of the encounter, a further series of analyses was performed on the data.

A series of multiple-regression analyses was carried out, in which the 60 subjects' ratings of the combined scenes for each of four dyads were used as dependent variables and their separate ratings of settings and communications were used as predictor variables. The unit of analysis was thus a combination of subjects and stimuli. This analysis was carried out for congruent ($N = 8$) and incongruent ($N = 8$) combinations of environmental and behavioral cues separately, and for all combinations together. This technique has been used by several researchers to analyze cue integration processes in social perception (Watson, 1972). Assuming the validity of the cue scale values used as predictors, these analyses provide a quantified representation of the respective contributions of the two sources of information, environment and nonverbal behavior, to judgments of the combined scenes. The results of these analyses are summarized in Table 1.

These results are congruent with the findings of the MANOVA analysis, indicating the generally greater weight of behavioral cues in the observers' ratings, with strong scale-specific differences, which suggests that some attributes of the encounter were considerably more affected by environmental cues than others. At the same time, it appears that the congruence or lack of congruence of environmental and behavioral cues had little effect on the observers' ratings of the combined scenes.

DISCUSSION

The present study, using specially developed and matched environmental and behavioral cues, was successful in demonstrating that the

Table 1. Multiple-regression analyses of judgments of combined scenes on separate ratings of settings and communications

	All combinations together			Congruent combinations			Incongruent combinations		
	Regression coefficient		Multiple	Regression coefficient		Multiple	Regression coefficient		Multiple
	NVC	Setting	R	NVC	Setting	R	NVC	Setting	R
Warm/cold	.50**	32**	.72	.54**	.35**	.72	.47**	.31**	.71
Simple/complex	.03	.21*	.24	–.03	.14	.16	.10	29**	.34
Active/passive	.11	–.01	.10	.04	.09	.09	.16	– 12	.19
Intimate/nonintimate	.69**	22**	.72	.65**	.24**	.69	.72**	.21*	.74
Pleasant/unpleasant	.51**	.34**	.72	.51**	.38**	.71	.50**	31**	.71
Tense/relaxed	44**	36**	70	.44**	.45**	72	.45**	.27*	.67

* $p < .01$. ** $p < .001$.

physical setting of an observed interaction contains important and nonobvious cues for observers. While the small number of stimuli judged and the small number of subjects sampled make generalizations difficult, the results are nevertheless interesting in pointing toward the important role physical environments play in social perception.

The first set of results demonstrates that both the environmental setting and the nonverbal behavioral communication of the interactors significantly affected the observers' judgments. The significant interaction of these two factors indicates some interesting cue-integration strategies.

The nature of this interaction is clearly represented in Figure 2. On all six scales, behavioral cues had more extreme values when displayed in setting 1 than in setting 2. Since the behavioral cues were the better predictors of judgments of combined scenes, it will be assumed that the interacting couples constituted the 'effective' stimulus and their environment the context. The reverse of this is equally possible, however. It appears that a setting judged to be comparatively warm, intimate, pleasant, and relaxed (setting 1, Figure 1) serves to accentuate interpersonal behaviors within it. On the other hand, a setting judged to be more cold, unintimate, unpleasant, and tense (setting 2, Figure 1) appears to attenuate behavioral cues within it, irrespective of whether interpersonal intimacy or lack of intimacy is expressed. From the evidence of this first analysis, it appears thus that environmental settings affect judgments of social encounters by providing information that increases or decreases the salience of the interpersonal behaviors observed within it.

It is also of direct interest to determine how congruent and incongruent combinations of environmental and behavioral cues interact. The second set of results, based on the multiple-regression analyses, is of relevance here. From Table 1 it appears that congruence or lack of congruence between the intimacy of the setting and that of the communication had no noticeable effect on the ratings of the combined scenes. This may be due to the fact that in only one of the two environments, setting 1, were nonverbal behaviors seen as sufficiently different.

The multiple-regression analyses also show some interesting scale-specific differences. On the whole, nonverbal behavior was more salient as a determinant of ratings than environment. Within this general conclusion, however, it would be possible to rank the different scales used here in terms of how sensitive they were to behavioral cues and environmental cues respectively (Table 1). Thus, in judgments of perceived *intimacy* of the encounters, behavioral cues weighted about three times more heavily than environmental cues. In judgments of the *warmth* and *pleasantness* of the encounters, behavioral cues still predominated but were only mar-

ginally more salient than environmental cues. On the *tense/relaxed* dimensions, behavioral and environmental cues were weighted equally in congruent combinations, while behavioral cues were about twice as important in incongruent combinations. Only on the relatively unimportant *simple/complex* scale were environmental cues consistently more heavily weighted than behavioral cues, while the *active/passive* scale turned out to be, on the whole, irrelevant to judgments of this encounter. These differences between scales in their sensitivity to behavioral and environmental cues is intuitively meaningful and represents a strong justification for the use of multiscale instruments in studies of social perception.

On the whole, the present results suggest that environmental cues play an important and nonobvious role in the perception of social episodes. While a much more extensive range of matched stimuli is needed before generalizations can be made, the findings are encouraging. Further studies of the effects of situational characteristics on social perception, using more realistic stimulus materials and more sophisticated, functional measurement methodologies, would be required to exactly determine the role of physical environmental cues in the perception of social episodes.

Notes

The authors gratefully acknowledge the suggestions and comments made by Michael Argyle. Requests for offprints should be sent to Joseph P. Forgas, School of Psychology, University of New South Wales, P.O. Box 1, Kensington 2033, N.S.W., Sydney, Australia. Received for publication July 13, 1976; revision, October 13, 1976.

References

Altmann, I., and Lett, E. E. 1970. The ecology of interpersonal relationships: A classification system and conceptual model. In *Social and psychological factors in stress,* ed. J. McGrath. New York: Holt, Rinehart and Winston.

Ball, D. W. 1970. An abortion clinic ethnography. In *Social psychology through symbolic interaction,* ed. G. P. Stone and H. A. Farberman. Waltham, Mass.: Ginn-Blaisdell.

Barker, R. F. 1968. *Ecological psychology.* Stanford, Calif.: Stanford University Press.

Barker, R. F., and Wright, H. F. 1955. *Midwest and its children.* Evanston, Ill.: Row, Peterson.

Bennett, D. J., and Bennett, J. D. 1970. Making the scene. In *Social psychology through symbolic interaction,* ed. G. P. Stone and H. A. Farberman. Waltham, Mass.: Ginn-Blaisdell.

Craik, K. H. 1973. Environmental psychology. *Annual Review of Psychology* 24:403–422.

Forgas, J. P. 1976. The perception of social episodes: Categorical and dimensional representations in two different social milieus. *Journal of Personality and Social Psychology* 34:199–209.

Goffman, E. 1959. *The presentation of self in everyday life*. Garden City, N.J.: Doubleday.

Goffman, E. 1963. *Behaviour in public places*. New York: Free Press.

Goodey, B. 1971. Perception of the environment. University of Birmingham Centre for Urban and Regional Studies, occasional paper 17.

Insel, P. M., and Moos, R. H. 1974. Psychological environments: Expanding the scope of human ecology. *American Psychologist* 29:179–188.

Mehrabian, A. 1968. The inference of attitudes from the posture, orientation, and distance of a communication. *Journal of Consulting and Clinical Psychology* 32:296–308.

Moos, R. H. 1973. Conceptualizations of human environments. *American Psychologist* 28:652–665.

Moos, R. H. 1974. *Evaluating treatment environments: A social ecological approach*. New York: Wiley.

Pace, C. R., and Stern, G. G. 1958. An approach to the measurement of psychological characteristics of college environments. *Journal of Educational Psychology* 49:269–277.

Proshansky, H. M., Ittelson, W. H., and Rivlin, L. G., eds. 1970. *Environmental psychology: Man and his physical setting*. New York: Holt, Rinehart and Winston.

Stern, G., Stein, M., and Bloom, B. 1956. *Methods in personality assessment* Glencoe, Ill.: Free Press.

Triandis, H. C. 1972. *The analysis of subjective culture*. New York: Wiley.

Warr, P. B., and Knapper, C. 1968. *The perception of people and events*. New York: Wiley.

Watson, G. S. 1972. Judgment of emotion from facial and contextual cue combinations. *Journal of Personality and Social Psychology* 24:334–342.

Wish, M., Deutsch, M., and Kaplan, S. J. 1976. Perceived dimensions of interpersonal relations. *Journal of Personality and Social Psychology* 33:409–420

AMERICAN JOURNAL OF PSYCHOLOGY
December 1977, Vol. 90, No. 4, pp. 645–654

An illusion of auditory saltation similar to the cutaneous "rabbit"

Christine D. Bremer, John B. Pittenger,
Rik Warren, and James J. Jenkins
University of Minnesota

Trains of brief clicks produced successively at 3 points in a horizontal array were not localized accurately. Observers reported clicks occurring in succession across the spaces between sources as well as at the sources themselves. The illusion is functionally related to interstimulus interval, number of clicks per speaker, and regularity of pulsing. It appears similar to Geldard and Sherrick's cutaneous "rabbit" illusion.

Geldard and Sherrick (1972) and Geldard (1975) have reported an illusion they call the cutaneous "rabbit." If, for example, 5 mechanical taps of equal duration and interstimulus interval are delivered successively to a point near the wrist, to the mid-forearm, and to a point near the elbow, the successive taps are not felt at these 3 loci only. While some of the taps are felt at these 3 points, others seem to be occurring at points between the contactors. In general, the taps are felt as a succession of uniformly spaced pulses in the region between the wrist and elbow. The experience is described as feeling like a tiny rabbit is hopping up one's arm.

A number of parametric characteristics of the cutaneous "rabbit" illusion are reported by Geldard and Sherrick. The number of contactors, distance between contactors, and the number of pulses per contactor do not appear to be crucial to the effect, although too many pulses per contactor can break down the effect. The illusion is, however, clearly related to the time between taps. With long interstimulus intervals (ISIs) the taps are veridically localized under the contactors. As the ISI decreases to about 200 msec, the perceived taps begin to "stray" noticeably from under the contractors, spreading out until at about 100 msec they seem uniformly distributed on a line connecting the contactors. The hopping effect is most vivid at about 50 msec. Irregularity in pulsing disturbs the effect. The effect is strongest when the interval between the last pulse at

one contactor and the first pulse at the next contactor is equal to the interval between pulses occurring at the same contactor. Lengthening the between-contactor ISI to 3 or 4 times the within-contactor ISI destroys the effect.

In the present study, we investigated an analogous auditory "rabbit," or auditory saltation illusion, and studied the influence of ISI, number of clicks per speaker, and regularity of pulsing on the illusion. Three related experiments were conducted, one with naïve and the others with knowledgeable observers.

EXPERIMENT I

METHOD

In the first experiment, 7 naïve observers were studied. They were volunteers from a general psychology course. The observers were tested individually in a small room lined with fiberglass batting to reduce sound reflection. The level of ambient noise in the room was approximately 50 dB SPL re 20 $\mu N/m^2$. Three speakers (Sound West Inc. Extraspeakers), chosen to produce similar-sounding clicks, were mounted in a curved horizontal array so that each speaker was 122 cm from the observer's head. The middle speaker was in the saggital plane of the observer and the end speakers were 24° to either side of the sagittal plane The 3 sound sources were masked from the observer's view by a thin black curtain that had a tape measure attached along it just above the speakers. The observer's head was positioned on a chin rest during the stimulus presentations.

The observers were asked to listen to series of sounds coming from behind the curtain. After each series, they were to make *X*s on numbered lines printed on response sheets. The *X*s were to correspond to the places behind the tape measure that they perceived as the locations of each of the sounds.

Observers were informed as to the number of sounds that would occur in each series. If 2 or more sounds were heard at the same place, the observer was to mark the corresponding *X*s one above the other on the response form.

Each click was of 20 msec duration. Though of relatively long duration (Geldard and Sherrick used taps 2 msec in duration), these clicks were perceived by observers as unitary and discrete. The speakers provided high damping such that successive signals were not interfered with. The speakers were driven by negative-going square waves of 9 volts. Each stimulus presentation consisted of a series of sounds starting in the left speaker and ending in the right speaker Forty-nine stimulus trains were presented, 22 of which were central to the experiment. For 18 trains, 3 clicks occurred in each speaker and the interval between successive clicks was constant within each train. ISIs (measured from click offset to click onset) varied from 10 msec to 250 msec. These stimuli allowed an appraisal of the effect of ISI on perceived location of the sounds. To study the effects of regularity of pulsing, 4 irregular series were presented. Three clicks occurred in each speaker, but the interval between clicks in adjacent speakers was approximately 3 times the interval between successive clicks within a speaker ISI for clicks within a speaker varied from 20 msec to 100 msec.

The remainder of the stimuli used other numbers of sounds per series, and some used only 2 of the speakers. These additional stimuli were included primarily to prevent the observers from responding in a stereotypic fashion to all trains. Each series was presented twice in succession before the observer made location judgments. The set of stimuli was presented in a different random order to each observer.

RESULTS

The written judgments of the observers of each series were categorized as 1, 2, or 3; "1" meaning that the sounds were reported at 3 locations only, 2 to 4 clicks at each point; "2" meaning that the clicks were reported as spread out beyond 3 points, but still in 3 distinct groups; and "3" meaning that the clicks were reported as being spread out more extensively, and in many cases as quite evenly distributed across a range of space.

Data from naïve subjects showed some variability. The modal responses of each subject to the 22 series are shown in Table 1. For regularly pulsed stimuli, 3 series were the basis for each mode; for irregular trains, 4 were the basis for the mode.

Table 1. Modal localization categories for naïve observers as a function of duration and regularity of ISI

Observer	Modal category						
	Regular trains (ISI interval, msec)						Irregular trains (ISI interval, msec)
	10, 15, 20	25, 30, 35	40, 45, 50	60, 70, 75	80, 90, 100	150, 200, 250	20, 30, 50, 100
1	3	2	*	1	1	1	1
2	3	2	3	1	1	1	1
3	3	2	2	2	*	1	1
4	3	2	2	3	2	2	1
5	3	2	2	2	*	1	**
6	1	*	1	*	1	1	1
7	2	2	2	2	2	2	2

* No unique mode; 1, 2, and 3 each occurred equally often.
** No unique mode; 1 and 2 each occurred equally often.

These results suggest an effect quite similar to the cutaneous "rabbit." All seven naïve observers made some category 3 responses, demonstrating the existence of an auditory "rabbit" or auditory saltation illusion.

All observers also made some category 1 responses. Most of these responses were made, as would be expected, to irregular stimulus trains or to regular trains having long ISIs. For 5 observers, the pattern of categories as a function of ISI is similar to that found for the cutaneous rabbit. Two observers do not show this pattern. One (#6) tends to localize most trains at 3 places, and another (#7) perceives most stimuli as 3 groups of slightly dispersed clicks. For both these observers, however, all 3 judgment categories occur as responses to regular trains. Irregular trains almost invariably received category 1 ratings at all ISIs.

The results obtained regarding the long-ISI regular trains indicate that the perception of saltation or spread in the short-ISI trains is not due simply to the regularity of spacing between the pulses, but is also dependent on ISI.

Other factors may, of course, be quite important, such as distance between speakers, arrangement of speakers in relation to the observer's head and characteristics of the pulses such as length, frequency spectrum, and amplitude. These factors remain to be investigated.

None of the observers described any of the stimuli as sounding like a hopping rabbit. (What, after all, does a hopping rabbit sound like?) The phenomenal experience was regularly described as sounding like a stick being run along a picket fence, or, especially at very short ISIs, a thumb along the edge of a comb.

EXPERIMENT II

In the second experiment, 2 observers were tested who were aware of the cutaneous-rabbit phenomenon and of the physical layout of the apparatus in the present study. The observers were staff members experienced in perceptual research. These observers heard the 22 trains of 9 clicks that were employed as critical trials for the naïve observers. Rather than marking *X*s, however, these observers simply categorized each stimulus series as 1, 2, or 3, according to the criteria used in scoring the naïve observers' responses in Experiment I.

The observers heard and judged the set of stimuli 12 times each. In each of 6 sessions, the set was presented twice with different random orders for each presentation, session, and observer.

RESULTS

The judgments of the knowledgeable observers show a pattern similar to that of the naïve observers, but with less variability. Figure 1 shows the frequency of judgment categories as a function of ISI for the regularly pulsed trains. Since there was no change in the pattern of responding over trials, the data from all 12 presentations of the set of stimuli are grouped. For both observers the trains with long ISIs are heard as well-localized. With ISIs between about 50 and 150 msec the pulses are judged

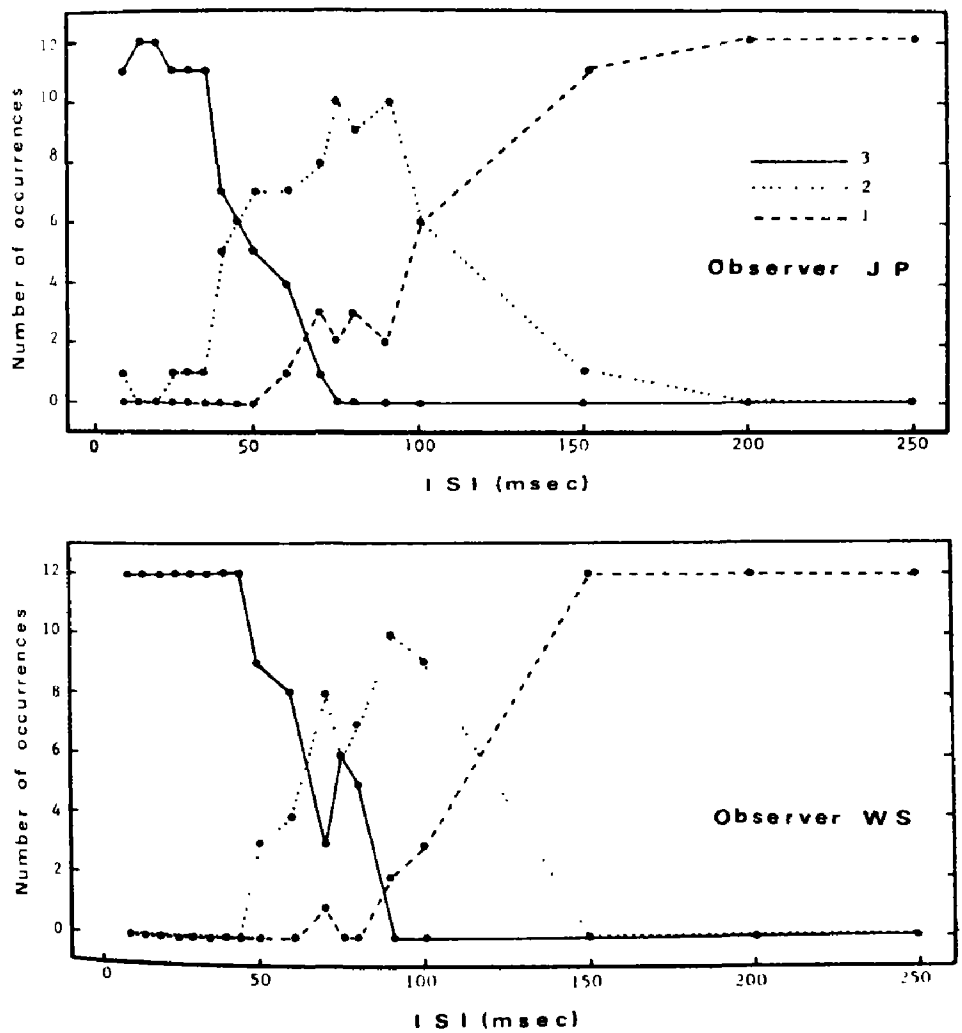


Figure 1. Distribution of category judgments of knowledgeable observers as a function of ISI

to be spread out from their actual loci and at ISIs shorter than 50 msec pulses seem to be evenly distributed along the entire line between the speakers. The 4 irregular trains were all well-localized with only a few reports of spread occurring for the most rapid of these.

EXPERIMENT III

The purpose of this experiment was to assess the effect of the number of clicks per speaker on the rabbit phenomenon. Two observers (University of Minnesota undergraduates) were tested who were knowledgeable about the phenomenon and were informed of the physical layout of the apparatus. Procedures in this experiment were similar to those in the second experiment, in which observers were asked to make category judgments of the rabbit effect. The major difference was that the number of clicks per speaker was systematically varied. There were either 2, 3, 4, or 5 clicks per speaker; the number of clicks per speaker was constant within a train. Thus the total number of clicks per train was 6, 9, 12, or 15. The 4 levels of clicks were crossed with 18 levels of ISI. The 72 possible combinations of clicks and ISIs constituted a trial block. Each observer was given 3 blocks, and the order of trains was randomly permuted within each block. The other ways in which the experiment differed from the second experiment were that the click length was reduced to 10 msec and there were no irregular trains.

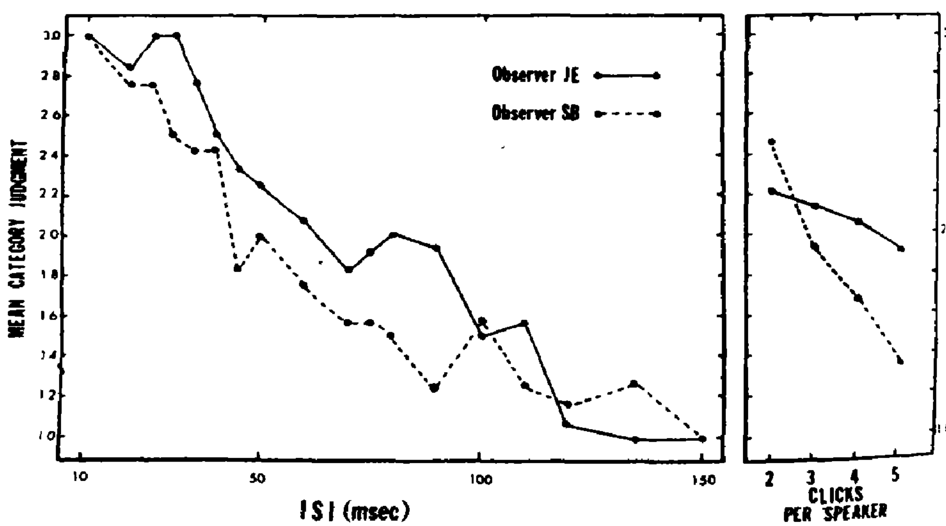


Figure 2. Mean of category judgments for the 2 observers in Experiment III as a function of ISI and clicks per speaker

RESULTS

The effect of ISI on the rabbit illusion was similar to that found in the second experiment: the longer the ISI, the more veridical the localization. The left side of Figure 2 shows the mean category judgment as a function of ISI. Here the data are collapsed over repetitions and number of clicks per speaker. When the data are collapsed over repetitions and ISI, as shown in the right side of Figure 2, it can be seen that the rabbit illusion is also a function of the number of clicks per speaker: the greater the number of clicks per speaker, the more veridical the localization.

For the following analyses, the mean category judgment of each observer at each of the 72 ISI × click combinations was paired with either the ISI or the number of clicks per speaker (thus $N = 72$ for each correlation). For observer JE, the linear correlation of mean category judgment with ISI was $-.93$; with the number of clicks per speaker, the correlation was $-.17$. For observer SB, the corresponding *r*s were $-.72$ and $-.52$, respectively.

Thus, using simple linear regression, ISI, by itself, accounts for 86% and 52% of the variance of the mean category judgments for observers JE and SB respectively. In contrast, the number of clicks per speaker accounts for 3% and 27% of the variance of observers JE and SB respectively. Since the proportion of variance accounted for by the number of clicks may have been depressed by the restricted range of clicks, and since the ranges of ISI and number of clicks were not equal, a direct comparison of the relative strengths of these factors in determining the rabbit effect is not justified. However, the proportion of variance accounted for by both factors is the simple sum of the individual effects since, by design, the correlation between ISI and clicks is zero. The resulting proportions of 89% and 79%, for observers JE and SB respectively, are impressive considering that only first-degree regression is involved.

Increasing the number of clicks per speaker and lengthening the within-speaker ISI increases the total time and activity at a single speaker. This greater time and activity at one place apparently aids localization. Geldard (1975), referring to the cutaneous rabbit, has noted that when the pulse train persists too long in one place, the taps tend to get "anchored" under the contactor. Although the relative contributions of ISI and number of pulses per location to the total "anchoring" effect remain to be further investigated, the results of the present study are suggestive.

A direct numerical analysis of the rabbit phenomenon as a function of total "dwell" time per speaker (or total train length) is difficult. This

is because the dwell time per speaker is a joint function of both ISI and the number (and duration) of clicks and is, therefore, substantially correlated with each. Dwell time is defined as the time from the first click's onset to the last click's offset. Since ISI here refers to the time between click offset and the next click onset, dwell time equals *click duration times the number of clicks* plus *ISI times the number of ISIs.* A further complication is the inequality of the ranges of ISIs and clicks used. However, a graphic analysis is illustrative of the effects of ISI, number of clicks, and dwell time per speaker on the rabbit effect.

Figure 3 was obtained by combining the results of both observers and by combining the results of adjacent values of ISI. Ratings for the following levels of ISI were pooled: 10, 20, 25, and 30; 35, 40, 45, and 50; 60, 70, 75, and 80; 90, 100, and 110; and 120, 135, and 150 msec. This pooling does not affect the primary trends of the data, but does help to better illustrate the effects of the 3 variables. Figure 3 presents iso-rating contours which were interpolated from the obtained mean ratings for the click × pooled ISI combinations. Veridicality of localization (i.e., lowered rating) is clearly seen to increase with increases in either the number of clicks per speaker or ISI.

The dotted lines in Figure 3 show 3 reference levels of equal dwell

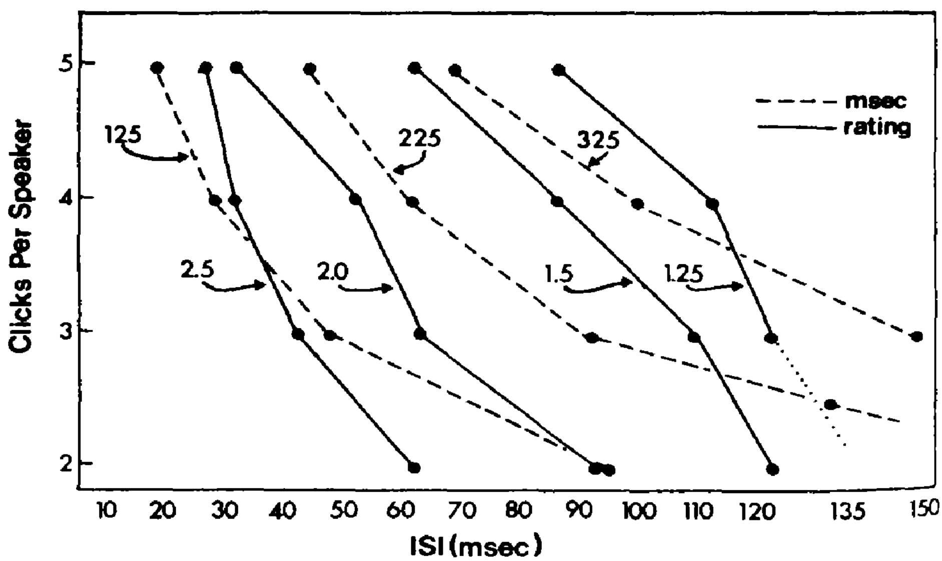


Figure 3. Iso-rating (solid lines) and iso-dwell time per speaker (dashed lines) contours as functions of ISI and clicks per speaker; higher ratings indicate a strong "rabbit" illusion, or conversely, poor localization. Pooled data from Experiment III

time per speaker. In general, the iso-rating contours are roughly parallel to the iso-dwell-time contours. The speaker dwell-time contours and the rating contours are most parallel (their curvatures are similar) at short speaker dwell times. This suggests that the relationship of dwell time and the rabbit effect is best at short dwell times: short dwell is generally associated with a strong rabbit effect. At longer dwell times, the iso-dwell-time contours and the iso-rating contours are least parallel (their curvatures appear opposite). Longer dwell times tend to be associated with more veridical localization, but the association appears weak and it would seem to be more profitable to account for good localization at long dwell times by looking directly at the ISI and click number levels.

DISCUSSION

Geldard and Sherrick (Note 1) have provided strong corroborating evidence for the existence of an auditory saltation illusion. Princeton University students enrolled in a laboratory course in perception conducted a study in which observers were asked to distinguish between illusory saltation (of the type we have described) and actual saltation (in which each pulse in the stimulus train is produced by a separate sound source).

Pairs of click patterns were presented to observers. Seven clicks occurred in each stimulus pattern. In each pair of patterns presented, one was produced by single clicks at each of 7 equally spaced speakers, and the other was produced by only the first, fourth, and seventh speakers in the array.

In a pair-comparison task, observers listened to 20 pairs of click patterns and were asked to judge which of each pair was the "rabbit" (illusory) as opposed to the "true" saltation.

The results obtained indicated that observers were less able to correctly choose the "rabbit" at shorter ISIs, and that the pattern of click distribution is important to the effect, i.e., more errors of discrimination occurred when the 7 clicks were relatively equally and symmetrically divided among the speakers (e.g., 2-3-2) than when the distribution was lopsided (e.g., 5-1-1). These experimenters concluded that the rabbit, or saltatory, effect does occur in appropriate circumstances in auditory space. Their results and our results are in accord with what one would expect were it the case that an auditory saltation illusion exists which is functionally similar to the cutaneous "rabbit" illusion.

In our auditory experiments and in those of the Princeton students, specific spatial and temporal parameters differed from those in the cutaneous case. However, the structural parameters were identical in that

there was in each case a linear array of stimulus sources pulsing sequentially. It is interesting, we think, that the cutaneous and auditory systems with all their differences in physical structure, respond in the same *inaccurate* manner to the same spatio-temporal *structure* of stimulation even though this stimulation is of 2 different types, i.e., mechanical contact and auditory stimulation. Further investigation of similarities and differences seem worthwhile.

Notes

This research was supported by grants to the University of Minnesota, Center for Research in Human Learning, from the National Science Foundation (GB 17590), the National Institute for Child Health and Human Development (HD 01136 and HD 00098), and the Graduate School of the University of Minnesota. We would like to thank Howard Graham for his assistance in conducting Experiments I and II. John B. Pittenger is now at the University of Arkansas, Little Rock, and Rik Warren is at State University of New York College at Buffalo. Requests for offprints should be sent to Christine D. Bremer, 205 Elliott Hall University of Minnesota, 75 East River Road, Minneapolis, MN 55455. Received for publication February 3, 1975; revision, January 6, 1977.

Reference note

1. Geldard, F. A., and Sherrick, C. E. *Princeton Cutaneous Research Project Report No. 24.* Princeton, N.J.: Princeton University, September, 1974.

References

Geldard, F. A. 1975. *Sensory Saltation: Metastability in the perceptual world.* Hillsdale, N.J.: Erlbaum.

Geldard, F. A., and Sherrick, C. E. 1972. *The cutaneous "rabbit": A perceptual illusion. Science* 178:178–179.

AMERICAN JOURNAL OF PSYCHOLOGY
December 1977, Vol. 90, No. 4, pp. 655–662

Discrete events in word encoding: The locus of elaboration

Donald A. Walter
University of Notre Dame

A model dealing with the function of elaboration in word encoding was evaluated using a 2-list recognition procedure that varied encoding time within the presentation list. The model predicted that elaboration, reflected in the incidence of false positives to associates of words presented in the recognition list, would increase as presentation time increased. No increase in false positives was found as presentation time lengthened, but an increase was found in the recognition of repeated words. Speculations were offered about a process in which activated elaborations of presented words were used to generate extremely specific encodings, facilitating recognition.

It is well known that if associates of words presented early in a continuous-recognition list are presented later, the number of false-positive responses to these associates will be greater than for neutral words in the same list. This has been shown by Underwood (1965), Kimble (1968), and Anisfeld and Knapp (1968).

Most authors have explained this phenomenon by an "elaboration" hypothesis of one sort or another. They collectively assert that the subject, in the process of encoding each individual word in a recognition list, elaborates the basic encoded representation of each in an effort to develop distinctive memory cues to be used for later recognition, and it is this elaboration process that leads to associative false-positive responses. Walter and Hellebusch (1974) have expanded on this hypothesis and imparted a temporal constraint to this encoding scheme. It consists of two parts, the first concerned with *comparison* and the second, with *elaboration* of the basic encoding of the word.

Their reason for this was twofold. First, in their study, retention interval did not affect the magnitude of the associative false-positive effect (the difference between associate and neutral word false-positive rates). Even with one intervening word between a root word (the first member

of an associated pair) and its associate, a substantial false-positive effect occurred. This meant that the distinction between the basic encoded representation and the elaborations of that representation was lost either immediately after or during encoding. It seemed to be more reasonable to posit an encoding scheme in which the differences between the basic encoded representation and its elaborations were obliterated during encoding, leading to immediate confusions about which was which. Second, the magnitude of associative false-positive effects was twice as large when the first member of the associative pair, the root word, was *correctly recognized* as a new word than when the root word was *misrecognized* as an old word. In the latter case, false-positive effects were not found, indicating that elaborations were somehow eliminated.

Walter and Hellebusch (1974) explained these results by hypothesizing a 2-stage word recognition sequence. The first stage involved a comparison of the encoding of the word presented with encodings of words previously presented. The second stage involved elaboration of the word encoding in an effort to better retain it in memory. Because elaboration occurs in the second stage, restriction of encoding time should decrease associative false positive effects in subsequent recognition as well as ordinary recognition scores.

There is already some evidence that this should be the case. It comes from experiments dealing with associative false-positive effects using intentional and incidental learning instructions. It seems that subjects more readily produce associative false positives under intentional instructions, but only when adequate encoding time is allowed. Both Wallace (1968) and Eagle and Ortof (1967) found robust false positives using presentation intervals greater than 1 sec. However, Cramer (1970a, 1970b) found that intentional instructions did not result in increased associative false-positive rates when encoding time was restricted to .5 sec. This evidence indicates that subjects need time to elaborate.

There is question, however, about the applicability of these studies to the present issue. First, it is unwise to compare results across experiments, and second, it is doubtful whether manipulation of encoding time in a between-subjects between-list design, where all words are presented for the same duration, is an effective way to minimize activities taking place in the latter stages of a temporal encoding process. It is possible that what may happen is that the entire encoding process may be accelerated to meet demands of a fast pace. In this case, results would not reveal the effects of an encoding process *interrupted in progress,* and thus would not reasonably reflect activities taking place in a second stage.

The present study uses a method which minimizes this problem. It

simply varies presentation duration as a within-subject variable, i.e., words presented in a list to be memorized appear for one of a number of different presentation durations. In this study, the durations used are .25, 15, 1, 2 sec. Presentation in this way prevents the subject from developing an expectation of the time a word will be available for encoding and thereby adjust his strategy (if any) to that expectation. Very short intervals will "cut into" the encoding process in progress, and long intervals will result in surplus encoding time. Since the procedure interrupts the process of encoding, the elaboration stage of the temporal sequence of encoding should suffer disproportionately in the case of short encoding durations.

METHOD

Design

The experiment consisted of a 2-list visual word recognition task in which words were presented in the first list and recognized in the second. Two independent variables were varied within subjects in a factorial design. Presentation duration (4 levels) was varied in List 1 presentation, and recognition item type was varied via List 2. Either repeats or associates of List 1 words were presented in the second list.

Subjects

Forty-eight University of Notre Dame undergraduates were self-selected from introductory classes and were given extra class credit for participation in the experiment. Twenty-four subjects participated in each of 2 nearly identical replications of the basic experiment.

Materials

Twenty-four word pairs were selected from Anisfeld and Knapp (1968). Each pair was made up of a root word and its associate. To each pair was added a control word that matched the associate with regard to animacy, concreteness, word class, frequency, and number of syllables, but was not an associate of the root word. The first member of each resulting triplet, the root word, was presented in List 1 of the recognition task. The associate and its control word were both presented in List 2 in nearby (±3) serial positions.

Another list of 20 words was drawn from the same distribution as the root words of the triplet list (the first member). These were presented in both List 1 and List 2. In addition, 40 filler words were drawn from the same distribution to minimize list-end effects and provide warm-up and practice sequences.

The first list of the 2-list experiment was made up of 64 words. Ten filler words were presented at the beginning of the list and 10 more at the end. Between these, 24 root words (from the root-associate-control triplets) and 20 first occurrences of repeat words were randomly positioned. Presentations of root words, repeat words, and filler words were evenly divided among the 4 presentation durations of .25, .5, 1, and 2 sec. The only changes made in the resulting list were to reposition adjacent repeated durations. This resulted in a list in

which word type and presentation duration were represented approximately equally in various sections of the list (except for filler words).

List 2 was a self-paced recognition list in which subjects made yes-no recognition responses. Four word types formed the list. Five filler words provided warm-up and were followed by 20 repeat words from List 1, 24 associates of List 1 root words, and 24 control words, one for each associate. All of the words (except the warm-up filler words) followed more or less the same order as their counterparts in List 1. Words appearing in the first part of List 1 appeared in the first part of List 2. Within 12-word groups the order was random. In all cases control words were positioned within 3 serial positions of their associative counterparts.

Procedure

The entire procedure was automated via a PDP8/I computer. Each subject was shown the equipment, which consisted of the computer, an oscilloscope display device, and a response device attached to an ASR-33 teletype keyboard. The response device had yes-no keys as well as 3 buttons labeled *probably, sure,* and *very sure*.

The display screen was programmed to display words in a 5 by 7 dot matrix for each letter. This resulted in a highly legible display. After the subject was familiarized with the equipment, a curtain was drawn to prevent the flashing lights on the computer console panel from distracting him. He was then handed a 4-page booklet containing the experimental instructions and told how to use it. The experimenter left the room, giving instructions to the subject to call him with any questions.

The first page of the booklet introduced the experiment and emphasized the importance of cooperation in following subsequent instructions carefully. On the second page the subject was told about the irregular presentation that was to occur in the practice trials and was cautioned not to memorize any of the words, but simply to get used to the display procedure. The third page contained instructions for the presentation list that would appear on the screen (List 1 of the 2-list experiment). The subject was instructed to memorize for later recognition each word in the list as it appeared, but not to spend any time thinking about a word when it was not physically on the screen. The fourth page contained instructions for the recognition list (List 2) of the experiment. The subject was told that a number of words would appear on the screen and that he was to make 2 responses to each. First, he was to determine whether or not the word shown appeared in the presentation list, and to press the yes or no keys to indicate his decision. It was emphasized that he should make his decision quickly and that if he felt that 3 sec had passed since the appearance of the word, he was to guess. Following the yes or no response, he was required to indicate his confidence in his answer by pressing one of the 3 confidence-rating keys. The pushing of any one of these initiated the next trial.

The instructions for each part of the experiment were read only after the completion of the previous part. The mean time for reading and comprehending instructions between the practice and presentation lists was 43 sec, between the presentation and recognition list, 56 sec. None of the subjects had any trouble following instructions except in the recognition phase (List 2). At this point 4 subjects of the 48 participating forgot that they had to press a confidence-rating

key to display the next word. They were instructed in the correct procedure and finished the experiment. This hiatus always occurred on the first filler word in List 2 and never resulted in an additional delay of more than 30 sec.

Twelve of the 24 subjects in each replication received both List 1 and List 2 conditions (but not the practice list) in a reversed order, except for filler words. This was done to minimize effects of the specific order of presentation durations in List 1. Each replication used a different quasi-random order of conditions in which the same presentation duration was never repeated.

The computer automatically counterbalanced the placement of specific words in list positions for both lists. For the root-associate-control triplets one complete counterbalancing occurred across the 24 subjects, one-half for subjects receiving the backward lists and one-half for subjects receiving the forward lists. Repeated words and filler words were also counterbalanced, but since only 20 of the former were used, the counterbalancing was slightly more than complete. The computer automatically recorded yes-no responses, confidence ratings, and latencies of the former.

RESULTS

The trends for forward and backward presentation conditions were virtually identical and their data were combined. Data for the 2 replications were combined for the same reason. Recognition responses for List 2 were of 3 types: hit responses for repeated words, pseudo hit responses for associates of List 1 words, and false positive responses for control words. In addition, latencies were available for either yes or no responses for all the confidence levels for each of the 3 word types. The mean proportions of old responses collapsed over presentation duration were .44 and .35 for associates and control words respectively. A test for differences between related sample means yielded a $t(23)$ of 4.02 ($p < .01$), showing that associates produced false positives significantly in excess of chance. Serial position had little effect on response bias. For successive quarters of List 2 mean control-word false-positive rates were $.35 \pm .02$.

The proportion of List 2 old (yes) responses for repeated words and associates is shown in Table 1 as a function of List 1 presentation duration. A 2-way repeated measures analysis-of-variance used the number of

Table 1. Proportion of old responses for recognition of associates and repeated words in List 2

	List 1 presentation time (in sec)			
	.25	50	1.00	2.00
Word type				
Repeated words (N = 240)	.60	.58	.78	.83
Associates (N = 288)	.42	.46	.44	.43

yes responses to both associates and repeated words as data. effects and the interaction showed significance. For List 2 $F(1, 47) = 76.09, p < .001$; for presentation duration, $F(3, 14$ $p < .005$; and for the interaction of both, $F(3, 141) = 3.23, p$

Since the false-positive rates for control words did not cha ciably as a function of serial position, these rates were taken indicator of the false-positive rate for computing the area of tl operator-characteristic (ROC) curves for both repeated word ciates as a function of presentation duration. This was done the areas formed by adjacent hit and false-positive rates for terion cutoff points defined by the 6 confidence ratings. A vis tion of the ROC curves formed in each case indicated that were approximately symmetrical about the negative diagonal square and met minimal assumptions for the signal-detection ı resulting areas are shown in Figure 1.

Unfortunately, some of the information needed to analyze of variance of ROC areas is consumed in the derivation of However, estimates of the standard errors of the area prop Figure 1) may be made. The standard error of the overall rep

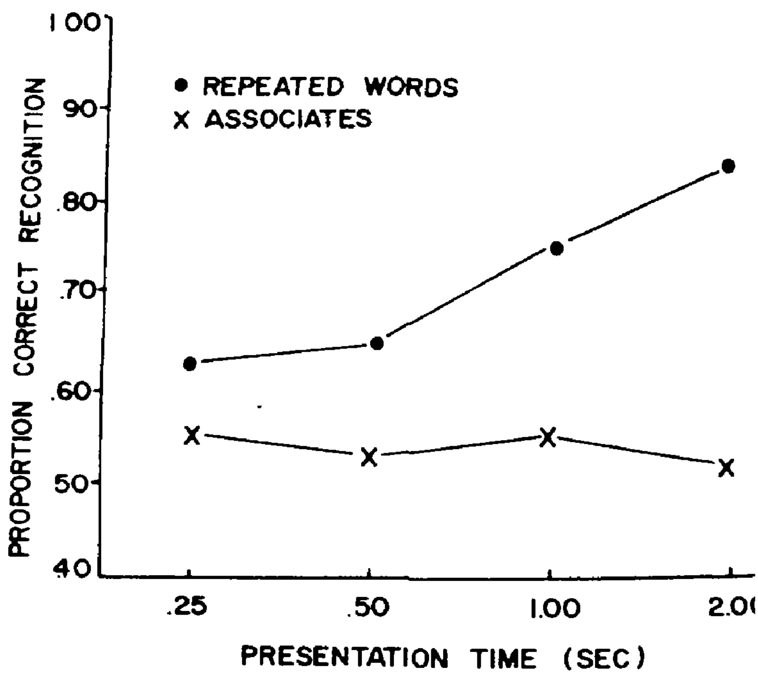


Figure 1. Proportion areas under the memory-operating characteris repeated words and associates as a function of 4 presentation durati

recognition area (.695) was .065. Thus, the difference between ROC area proportions for words presented for .25 and 2 sec was greater than 3 standard errors, certainly a significant monotonic trend. On the other hand, all of the area of recognition proportions for associates were well within one-half of a standard error (.07) of the overall mean of the four (.54), leaving no reason to hypothesize a real difference between them.

DISCUSSION

Specific predictions of a model of word encoding proposed by Walter and Hellebusch (1974) were tested: specifically, that an elaboration stage followed a comparison (with memory) stage, and that the shortening of encoding time to relatively severe durations should tend to "lop off" the second stage, inhibiting associative elaboration. This did not happen. No decrease in false-positive effects was found in the case of shortened encoding intervals.

What is happening during encoding may be inferred from these solid findings. First, it is virtually certain that *elaboration* occurs in recognition encoding (except for incidental learning situations and situations in which the word to be encoded is thought to be repeated). Second, elaboration occurs *early* in the encoding process. Third, *recognition scores increase* without any measurable increase in elaboration.

It is clear that the subject engages in a type of associative elaboration early in the encoding process that is reflected by false positives, but that such elaboration is not beneficial to later recognition. This apparent anomaly may be explained by assuming that the subject uses the latter portions of the presentation interval to *select* and refine memory representations activated as a result of early elaboration in such a way as to generate memory representations that might enable him to successfully discriminate the encoded word later in the recognition task. This selection process, however, does not take the form of simply "weeding-out" elaborations represented by the associative false-positive effects. It is likely that the process involved here may include reduction to a more specific level and idiosyncratic level, which is not probed by the associative false-positive effects at all.

Notes

Donald A. Walter is now at the West Virginia College of Graduate Studies, Institute, WV 25111, where requests for offprints should be addressed to him.

References

Anisfeld, M., and Knapp, M. 1968. Association, synonymity and directionality in false recognition. *Journal of Experimental Psychology* 77:171–178.

Cramer, P. 1970a. Magnitude and selectivity as independent factors in semantic generalization. *Journal of Verbal Learning and Verbal Behavior* 9:509–524.

Cramer, P. 1970b. Semantic generalization: IAR locus and instructions. *Journal of Experimental Psychology* 83:266–273.

Eagle, M., and Ortof, E. 1967. The level of attention upon "phonetic" recognition errors, *Journal of Verbal Learning and Verbal Behavior* 6:226–231.

Kimble, G. A. 1968. Mediating associations. *Journal of Experimental Psychology* 76:263–266.

Underwood, B. J. 1965. False recognition produced by implicit verbal responses, *Journal of Experimental Psychology* 70:122–129.

Wallace, W. P. 1968. Incidental learning: The influence of associative similarity in producing false recognition. *Journal of Verbal Learning and Verbal Behavior* 7:50–54.

Walter, D., and Hellebusch, S. 1974. The role of associative elaboration in word recognition: Evidence for a two-stage test elaboration sequence. *Journal of Experimental Psychology* 103:79–84.

AMERICAN JOURNAL OF PSYCHOLOGY
December 1977, Vol. 90, No. 4, pp. 663–674

Partial reinforcement and the learning of a strategy in the rat

Paul T. P. Wong
Trent University, Canada

When reinforcement was contingent on response variation, rats exhibited spontaneous stop-and-look behavior and acquired the strategy of response variation. This strategy-learning paradigm constituted a skill partial-reinforcement (PRF) condition where reinforcement was both predictable and controllable. Skill PRF produced greater resistance to extinction as compared with either a chance PRF control group yoked for percentage and pattern of reinforcement, or a variation consistent reinforcement control yoked for frequency of response variation. The result was interpreted in terms of Weiner's attribution model.

Estes (1971) has stated that behavior in infrahuman animals is best accountable in terms of specific stimulus-response (SR) sequences, while instrumental behavior of mature humans is best understood in terms of rules, principles, and strategies. Most of the research findings in animal learning do seem to justify Estes's discontinuity position. However, it may be argued that due to the prevailing influence of reflexology in animal learning research, the demand characteristics of testing conditions and the dependent variables employed tend to produce results that favor an SR analysis. It has been suggested that less-restrictive testing situations may bring forth evidence of strategy or rule learning in infrahuman animals (Wong, 1975; 1977a).

According to Gagné (1965), rule learning involves learning such propositions as "If *A* then *B*," where *A* and *B* are concepts. Scandura (1970, p. 519) has pointed out that "a rule can be denoted by a function whose domain is a set of stimuli and whose range is a set of responses." Skinner (1969) conceptualized rules as derived from contingencies of reinforcement in the form of formulations or descriptions that specify the occasions, responses, and their consequences. Regardless of these differences in definitions, it is generally agreed that rules are higher-order routines capable of generating a set of responses with common properties and

appropriate to certain situations. Rules are often termed strategies to denote that there are different means, with varying degrees of efficiency, to solve a particular problem. In the present context, strategy and rule constructs were used interchangeably.

Several types of strategies have been described in the animal-learning literature. A "win-stay, lose-shift" strategy may be inferred in repeated reversal learning and various types of learning-set tasks (e.g., Goodnow and Pettigrew, 1955; Harlow, 1949; Schrier, 1974). Trans-response and trans-situational transfer of the partial reinforcement extinction effect have been attributed to the acquisition of a "try" strategy (McCuller, Wong, and Amsel, 1976; Wong, 1977b; Wong and Amsel, 1976). A "choice-point strategy" in which the animals stop before a choice point and scan both the positive and negative stimuli has been observed in discrimination learning (Hall, 1973; Mandler and Hooper, 1967). The above are instances of *implicit* or indirect strategy learning, because reinforcement is not explicitly contingent on the utilization of these strategies. Their emergence is probably spontaneous and has the effect of facilitating goal attainment.

It is likely that exposure to any contingency of reinforcement may result in some kind of implicit strategy learning. In a Skinner box, when a rat arrives at the appropriate behavior of lever pressing through direct contact with the contingency of reinforcement, the behavior is said to be contingency-shaped. In this situation, the animal may even construct a broad, higher-order rule that if you work, you shall receive your reward (cf., Skinner, 1969, pp. 143–45). The extraction of this rule should have profound and pervasive effects on the animal's subsequent interaction with its environment. The knoweldge that the world is predictable and controllable and that work leads to goal attainment should predispose the animal to operate on its environment whenever in need. This hypothesis can be readily tested. Unfortunately, animal-learning psychologists rarely look beyond specific criterion responses to search for evidence of implicit rule learning.

Explicit rule learning refers to conditions where reinforcement is contingent on the acquisition and utilization of rules. Pryor, Haag, and O'Reilly (1969) reinforced porpoises to emit novel behaviors that had not been previously reinforced. This is an example of explicit rule learning where reinforcement was contingent on behavior that conformed to a predetermined relational rule that may be formulated as, "If response is different from preceding responses, then the outcome is food reward." This generative-response class of novel behavior appeared to be rule-governed, because it consisted of only topographies that had not been

previously reinforced. What was learned by these porpoises could not possibly be conditioned reflexes but could be more appropriately conceptualized as a rule that generated many instances of the same response class, all different in response topography. In this case, although novel behaviors are said to be rule-governed, the rule was acquired through prolonged direct contact with the contingency of reinforcement.

In view of the scarcity of evidence of both implicit and explicit strategy learning in rodents, the present study represents an initial step toward filling this knowledge gap. The main purpose of the present experiment was then to demonstrate the rat's capability to learn a response variation strategy through explicit reinforcement. Since the behavioral repertoire of rats is much more limited than that of porpoises, a complex maze that permitted an infinite variety of routes leading to the goalbox was used. Reinforcement was contingent on a relational rule that may be stated as: "If response (i.e., the route) is different from the immediately preceding one, then reinforcement."[1]

In the present situation, reinforcement was not contingent on a specific response topography but on its *relation* to the response topography on the last trial. This treatment constituted *skill* partial reinforcement (PRF) because only a subset of instrumental responses conforming to the predetermined relational rule was reinforced. Further, reinforcement was both predictable and controllable, and percentage of reinforcement obtained depended on how well the subject learned this relational rule.

It has been questioned whether the partial reinforcement effect (PRE) can be obtained under skill conditions (Feather, 1962). Previous skill vs. chance experiments with humans as subjects have failed to demonstrate the PRE under the skill condition when it was defined by the instructional set that success depended on skill, or persistence was solely defined by subjective expectancy of success (Altshuler and Kassinore, 1975; James and Rotter, 1958; Rotter, Liverant, and Crowne, 1961). However, when the skill condition was defined by contingency of reinforcement and persistence was measured by performance, some evidence of the PRE has been obtained in human subjects (DiCiaula, Martin, and Lotsof, 1968; Wong, Note 1). Further research is needed to directly compare instruction-defined and contingency-defined skill conditions, and to determine how persistence in performance is related to subjective expectancy of success as well as other cognitive factors.

In the animal-learning literature, discrimination training approximates a skill PRF treatment. Jenkins (1961) pointed out that in the early stage of discrimination, subjects responded to both S+ and S−, and were therefore only partially reinforced. It should also be noted that in dis-

crimination learning, percentage of reinforcement depends on discriminative skill and the subject essentially experiences continuous reinforcement (CRF) when it stops responding to S−. Two papers have reported greater resistance to extinction to S+ following discriminative operant training than following nondiscriminative CRF training (D'Amato, Schiff, and Jagoda, 1962; Jenkins, 1961), while one reported no difference (Birch, Allison, and House, 1963). This discrepancy may be due to the different levels of discrimination learning achieved prior to extinction to S+, because these differences should result in different amounts of generalized inhibitory tendency previously accrued to S−. In the present strategy-learning paradigm, learning the skill of response variation does not depend on inhibitory learning of not responding to S−; therefore extinction performance should not be confounded by generalized inhibition. The second purpose of the present experiment was to compare the effects of skill and chance PRF training on resistance to extinction.

The skill PRF condition was defined by the variation strategy learning treatment (V-PRF) in which reinforcement was contingent on Response 2 being different from Response 1. The appropriate CRF control for the V-PRF treatment was a variation-consistent reinforcement group (V-CRF) that received consistent reinforcement but was yoked with V-PRF for frequency of response variation. (In V-PRF, variation was imposed by reinforcement contingency; in V-CRF, variation was imposed by blocking individual routes.) To compare skill and chance conditions of PRF, a yoked PRF control was included which received the same percentage and pattern of reinforcement as V-PRF subjects, but reinforcement was not contingent on response variation. A conventional CRF group was also included to establish the conventional PRE (in comparison with the PRF group) and to assess the effect of response variation apart from PRF on extinction (in comparison with the V-CRF group). These 4 treatments may be treated as a 2 × 2 factorial design with 2 levels of Reinforcement (CRF vs. PRF) and 2 levels of Variation (Presence vs. Absence of imposed response variation).

METHOD

Subjects

Subjects were 28 male albino rats of the Sprague Dawley strain, approximately 90 days old. Before experimentation, they were handled for 2 weeks and reduced to 80% of their ad lib feeding body weights. They were then maintained at 80% body weights with water always available throughout the experiment.

Apparatus

The maze was made of unpainted plywood 1.89 cm thick and measured 120 cm long by 30 cm wide by 23 cm high. In the center of each end wall was an opening 12 cm wide by 23 cm high leading to an endbox 14 cm deep, made of plywood and separated by a Plexiglas guillotine door. One endbox served as startbox, the other as feeding box. Three parallel wood partitions .64 cm thick, measuring 30 by 30 cm, when inserted into the maze 30 cm apart, divided the maze into 4 equal compartments. Each partition had 2 openings 8 cm in diameter, with the center of each diameter 7 cm away from the maze floor and the nearest sidewall. One neutral set of partitions was painted grey. One set of partitions was painted black on the left and white on the right, split on the midline. An aluminum guillotine door lowered from behind the third partition separated the fourth compartment as a goalbox. The startbox, goalbox, and feeding box were covered with Plexiglas lids.

Procedure

Pretraining. This phase was designed to train subjects to go through the openings of the neutral set of partitions.

On Day 1, the rats were given 2 trials of goalbox placement with an intertrial interval (ITI) of about 3 hours. On each trial, subject was allowed 30 sec of access to a dish of wet mash placed at the entrance of the feeding box. Failing to eat in 3 min, subject was removed and given an additional opportunity to consume wet mash from an identical dish in the home cage.

On Days 2 to 3, the rats received 2 trials per day; Days 4 to 5, 4 trials per day; Days 6 to 7, 8 trials per day; Days 8 to 9, 10 trials per day. From Days 2 to 5, only the third partition was inserted. From Days 6 to 9, all 3 partitions were inserted. On each of these 8 days, trials were massed, reinforcement was 10 sec of access to wet mash, and subject was gently guided into the goalbox if it failed to enter it in 3 min. At the end of pretraining, all subjects were able to hurdle-jump through the openings of the 3 partitions in less than 1 min.

Acquisition. Subjects received 10 massed trials per day for 25 days in the maze with the 3 black-and-white partitions inserted. On each trial, subject was gently prodded when it failed to enter the goalbox in 1 min. On rewarded (R) trials, the feeding-box guillotine door was lifted as soon as the rat approached it, to permit 10 sec of access to wet mash. On nonrewarded (N) trials, the dish of wet mash was not accessible but was visible behind the closed feeding-box guillotine door, and subject was confined in the goalbox for 20 sec.

Subjects were tested in squads of 4, one from each of the 4 treatment conditions. On each day, the running sequence of these squads was randomized. Within each squad, the order was randomized with the restriction that V-PRF subject was run before its yoked PRF and V-CRF mates.

V-PRF subjects were reinforced consistently on the first trial of each day. This was intended to maintain responding when subjects failed to vary their routes for many consecutive trials. In the remaining 9 trials of each testing day, V-PRF subjects were reinforced when their route varied from the immediately preceding one. In each block of 10 trials, they could receive 100% reinforcement if they made 9 consecutive response variations.

Each PRF subject was yoked with one V-PRF subject in terms of frequency and pattern of reinforcement. Whenever a V-PRF subject earned a reinforcement, its yoked PRF mate was also reinforced on that trial regardless of whether it varied its response.

Each V-CRF subject was yoked with one V-PRF subject in terms of frequency of response variation. Whenever a V-PRF subject varied its route, its yoked V-CRF mate was forced to vary its response by having the entrance to its preceding route blocked by an aluminum door. However, blocking was not applied whenever the V-CRF yoked mate had previously varied the response on its own accord. Both V-CRF and CRF subjects were consistently reinforced.

Extinction. All the rats received 5 massed trials per day for 8 consecutive days. Running order of squads and subjects within each squad was randomized on each day. If subject entered the goalbox within the allotted time of 1 min, it was confined in the goalbox for 20 sec with wet mash inaccessible but visible behind the feeding-box guillotine door. If subject failed to enter the goalbox in 1 min, it was picked up, held for 10 sec (the amount of time typically required to complete the data sheet and transport rats from goalbox to startbox), and then returned to the startbox. If the rat escaped by jumping onto the side wall or partition, it was picked up, held for 10 sec and placed in the startbox to start another trial.

RESULTS AND DISCUSSION

Acquisition. Various acquisition measures are summarized in Table 1. It is clear from the first column that reinforcement of response variation was very effective. V-PRF subjects made significantly more response variations than PRF subjects ($t = 16.36$, $p < .001$). All but one V-PRF sub-

Table 1. Mean frequency data in acquisition

	1	2	3	4	5	6	7
Groups	Response variation[a]	Total N trials[b]	Variation after N	Total R trials[b]	Variation after R	Stop	Stop-look
V-PRF	149.28	70 86	39 14	154.14	110.14	49.64	53 39
PRF	15.57	70.86	7.57	154.14	8	16.39	11 85
V-CRF	149.28	—	—	225	149.28	36.57	19.46
CRF	5 75	—	—	225	5.75	14.89	4.46

[a] In each block of 10 trials, the possible number of variations from the preceding trial is 9. Therefore, considering only within-block variations, the total maximum number of response variations in 25 blocks of acquisition trials is 225 for each subject.

[b] Total N trials + Total R trials = 225. The last trial of each block (whether N or R) was not included in this total, because response variation from the last trial could occur only on the following day of testing and was not included in the response variation data.

ject at different points of acquisition reached the criterion of perfect performance (i.e., 9 variations in a block of 10 trials). Even the imperfect subject reached the level of 7 variations per block on 3 occasions. However, no subject was able to maintain the perfect criterion over several blocks. The average percentage of reinforcement was 70%. Consistent with previous findings (e.g., Wong, 1977), response variation was greater under PRF than CRF condition ($t = 2.08, p < .05$), but none of the PRF subjects ever made more than 6 variations per block.

It is also instructive to compare the probability of response variations after N trials with that after R trials (see columns 2 to 5). In view of previous observations of the "win-stay, lose-shift" strategy, it was expected that the probability of response variation was greater after N trials than after R trials in the PRF condition. This prediction was confirmed, with the percentage of variation after N trials (10.9%) significantly higher than after R trials (4.7%), $t = 9.16, p < .001$. If the V-PRF subjects also adopted a "win-stay, lose-shift" strategy, they would be reinforced only every other trial, because the consequence of repeating a just previously reinforced response was nonreward. To maximize reinforcement, they had to shift after each reinforcement. Therefore, the probability of shifting should be greater after R trials for V-PRF subjects. Results supported this prediction, with the percentage of variation after R trials (70.6%) significantly higher than after N trials (56.3%), $t = 2.59, p < .05$.

In addition to the above finding of strategy learning shaped by an explicit contingency of reinforcement, there is also evidence of implicit strategy utilization not directly related to reinforcement. Columns 6 and 7 show respectively mean numbers of trials in which the rats paused briefly before the choice point (Stop) and paused to inspect the left and right entry points (Stop-Look). Interexperimenter agreement in both measures was higher than 80%. Analysis of variance (Anova) of Stop data yielded a significant main effect of Reinforcement [$F(1, 24) = 4.94$, $p < .05$] and a significant main effect of Variation [$F(1, 24) = 70.27$, $p < .001$]. Anova of Stop-Look data revealed significant effects of Reinforcement [$F(1, 24) = 24.06, p < .001$], Variation [$F(1, 24) = 45.07, p < .001$], and their interaction [$F(1, 24) = 9.93, p < .01$]. This interaction was due mainly to a greater difference between V-PRF and V-CRF ($t = 4.23, p < .01$) than between PRF and CRF ($t = 2.85, p < .05$). In other pairwise comparisons, V-PRF was significantly different from PRF ($t = 5.50, p < .001$) and V-CRF from CRF ($t = 5.20, p < .001$).

The significantly higher level of Stop and Stop-Look by V-CRF relative to CRF group was due mainly to occasional route blocking in the

former. When the entrance to a preferred, dominant route was blocked, it was necessary for the subject to stop and look for an alternative. The greater frequency of Stop-Look by PRF groups corroborates previous findings of greater investigating activities in PRF condition, e.g., sniffing (Wong, in press, a) and hole-exploration (Wong, in press).

Of special interest was the higher level of Stop and Stop-Look in V-PRF condition relative to both V-CRF and PRF groups. Rotter (1966) reported that when human subjects were informed that performance outcome depended on their own behavior rather than on chance, they paid more attention to external cues that might facilitate performance. In the present study, rats learned through the contingency of reinforcement that performance outcome depended on their own behavior, and this learning about contingency should increase the probability of attending to external cues that might facilitate goal attainment. This Stop-Look information-processing strategy is similar to the "choice-point" strategy observed in discrimination learning, where performance outcome also depends on making the correct response (Mandler and Hooper, 1967).

More recently, the above acquisition findings have been replicated in hooded rats, but attempts to reinforce the response variation strategy in hippocampally damaged rats have not been successful (Wong and Winocur, Note 2). It is noteworthy that these hippocampal rats also rarely exhibited the Stop-Look information-seeking strategy.

Extinction. Various aspects of extinction behavior are summarized in Table 2. Anovas were performed on all these measures. Where a signifi-

Table 2. Mean frequency data in extinction

Groups	1 Trials to 1-min criterion	2 Total no. of criterion. trials	3 Trials to retrace	4 Total no. of retracing	5 Trials to escape	6 Total no. of escapes	7 Total no. of eliminations	8 Response variation[a]
V-PRF	23.71	4 85	38.00	0.42	39.43	0.07	3.00	23 43
PRF	9.00	13.00	32.28	0.57	32.14	0.81	1.78	3 29
V-CRF	5.14	18.85	5.85	4.85	16.43	11 71	0.42	6 86
CRF	6.28	16.71	15.28	5.14	28.43	6.25	0.85	3 71

[a] In each block of 5 trials, the possible number of variations from the preceding completed trial is 4. (Trials were considered completed if subject traversed the maze and entered the goalbox.) The maximum number of response variations in 8 blocks of extinction trials is 32 for each subject.

cant Reinforcement × Variation interaction was obtained, only significant t values of planned pairwise comparisons were reported (i.e., V-PRF vs. V-CRF, PRF vs. CRF, V-CRF vs. CRF and V-PRF vs. PRF).

In both trials to 1-min criterion and total number of 1-min criterion trials (Columns 1 and 2), a significant Reinforcement × Variation interaction was obtained, $Fs(1, 24) = 8.29, 7.57$, respectively, $p < .05$ in both cases. In both of these measures, V-PRF group was more resistant to extinction than PRF [$ts = 3.04, 2.92, p < .05$ in both cases] as well as V-CRF [$ts = 5.85, 15.65, p < .001$ in both cases], but there was no reliable difference between PRF and CRF. Trials to retrace and total number of trials where retracing occurred (Columns 3 and 4) reflect the tendency to avoid the goalbox, because the rats were considered retracing only when they turned back before the goalbox door. When a subject never retraced throughout extinction, a score of 40, the total number of extinction trials, was accorded. In these two measures, CRF groups exhibited a greater avoidance tendency, $Fs(1, 24) = 34.48, 9.65$; $p < .01$, respectively.

In trials to escape (Column 5), the significant Reinforcement × Variation interaction [$F(1, 24) = 6.06, p < .05$] was due to a significant difference between V-PRF and V-CRF [$t = 2.37, p < .05$] and the absence of a significant difference between PRF and CRF.

In terms of total number of trials on which escape occurred (Column 6), there was also a significant Reinforcement × Variation interaction [$F(1, 24) = 15.29, p < .001$]. V-PRF group escaped less frequently than V-CRF group [$t = 6.82, p < .001$], but there was no difference between PRF and CRF groups. It is also interesting to note that while V-PRF and PRF groups did not differ, V-CRF group escaped more frequently than CRF group [$t = 3.74, p < .01$].

Another aspect of emotional reaction to extinction was elimination (Column 7). Reinforcement main effect approached significance, $F(1, 24) = 3.75, p = 0.65$, suggesting that the more persistent approach to the empty goalbox, the more emotional eliminations.

In response variation (Column 8), both the Reinforcement and Variation main effects were significant, $Fs(1, 24) = 61.69, 128.34$, respectively, $p < .001$ in both cases.

In extinction the Stop and Stop-Look measures were confounded and not recorded, because retracing and escape necessarily involved either Stop or Stop-Look.

In summary, V-PRF group was more persistent than PRF group in all extinction measures (Columns 1 to 6), although this difference was significant only in trials to criterion and number of criterion trials. Since both groups had the same percentage and pattern of reinforcement, this differ-

ence was puzzling. One interpretative possibility is that greater response variation in the V-PRF condition should require more trials to extinguish (e.g., Newberry, 1971). This interpretation is weakened by the finding that although the V-CRF group also experienced more response variation, and even exhibited more response variation than CRF subjects during extinction, there was no evidence of greater persistence by V-CRF subjects in any of the extinction measures. Another hypothesis is that under the contingency-defined skill PRF condition, subjects may "try" harder and may be more apt to attribute performance outcome to effort than under the chance condition. According to Weiner's model, effort attribution should mediate persistence (Weiner, Frieze, Kukla, Reed, Rest, and Rosenbaum, 1972). There is already some evidence that effort attribution is positively related to behavioral persistence during extinction (Dimitroff, Note 3; Wong, Note 1) and expectancy of success (Weiner, Heckhausen, Meyer, and Cook, 1972). These findings suggest that resistance to extinction is mainly determined by effort attribution. When subjects believe that reinforcement depends on their effort, they will try harder (i.e., persist longer) during extinction. This attributional analysis would predict that the skill PRF training should produce greater persistence than the chance PRF training, only when the skill is not mastered. Once subjects have mastered the skill and are consistently reinforced, they are more likely to change from an effort attribution (which is internal and unstable) to an ability attribution (which is internal and stable) than their yoked chance controls. Following mastery of the task the skill group should be less persistent than the chance controls because the former should be more likely to make an ability attribution. That is, they are more likely to perceive extinction as a new task for which they are deficient in ability. This hypothesis remains to be tested.

In conclusion, the present experiment has demonstrated the acquisition of a response variation strategy by the rat and the accompanying emergence of a Stop-Look information-seeking strategy. Another major finding is that V-PRF was more persistent than V-CRF, thus providing the first clear evidence of a skill PRE in the rat. V-PRF group was also more persistent than the yoked chance PRF control in some extinction measures; to account for this finding, it may not be unreasonable to apply an attribution analysis to the lowly rodents.

Notes

This research was supported by NRC Grant No. A8635. Thanks are due to Janet Olds and Wanda Grife for collecting the data. Requests for reprints should be

addressed to Paul T. P. Wong, Department of Psychology, Trent University, Peterborough, Ontario, K9J 7B8, Canada. Received for publication October 2, 1975; revision, August 17, 1976.

1. Since the maze consisted of 3 sets of binary choice points and the rats were allowed to retrace and repeat at any choice point, an infinite number of different routes were possible. Reinforcement was contingent on within-day and not on between-days response variation. In the latter case, subjects would have to remember the specific route and its consequence on the last trial of the preceding day. Previous research (cf., Surridge and Amsel, 1965) has questioned whether rats can remember goal events with a 24-hour intertrial-interval.

Reference notes

1. Wong, P. T. P. Developing productive persistence in conceptual problem solving: The constructive role of failure in the teaching-learning process. Final report submitted to the Ontario Ministry of Education, 1976.
2. Wong, P. T. P., and Winocur, G. Strategy learning in hippocampectomized rats. Unpublished study, Trent University, Peterborough, Ontario, 1976.
3. Dimitroff, G. Attribution theory and the partial reinforcement effect. M.A. thesis, University of Toronto, 1976.

References

Altshuler, R., and Kassinore, H. 1975. The effects of skill and chance instructional sets, schedule of reinforcement, and sex on children's temporal persistence. *Child Development* 46:258–262.

Birch, D., Allison, J. K., and House, R. F. 1963. Extinction performance following discrimination training. *Journal of Experimental Psychology* 65:148–155.

D'Amato, M. R., Schiff, D., and Jagoda, H. 1962. Resistance to extinction after varying amounts of discriminative or nondiscriminative instrumental training. *Journal of Experimental Psychology* 64:526–532.

DiCiaula, P. J., Jr., Martin, R. B., and Lotsof, E. J. 1968. Persistence and intermittent reinforcement. *Psychological Reports* 23:739–742.

Estes, W. K. 1971. Reward in human learning: Theoretical issues and strategic choice points. In *The nature of reinforcement,* ed. R. Glazer. New York: Academic Press, pp. 16–44.

Feather, N. T. 1962. The study of persistence. *Psychological Bulletin* 59:94–115.

Gagné, R. M. 1965. *The conditions of learning.* New York: Holt, Rinehart, and Winston.

Goodnow, J. J., and Pettigrew, T. F. 1955. Effect of prior patterns on experience upon strategies and learning sets. *Journal of Experimental Psychology* 49: 381–389.

Hall, G. 1973. Response strategies after overtraining in the jumping stand. *Animal Learning and Behavior* 1:157–160.

Harlow, H. F. 1949. The formation of learning sets. *Psychological Review* 56: 51–65.

James, W. H., and Rotter, J. B. 1958. Partial and 100% reinforcement under chance and skill conditions. *Journal of Experimental Psychology* 55:397–403.

Jenkins, H. M. 1961. The effect of discrimination training on extinction. *Journal of Experimental Psychology* 61:111–121.

Mandler, J. M., and Hooper, W. R. 1967. Overtraining and goal approach strategies in discrimination reversal. *Quarterly Journal of Experimental Psychology* 19:142–149.

McCuller, T., Wong, P. T. P., and Amsel, A. 1976. Transfer of persistence from fixed-ratio barpressing training to runway extinction. *Animal Learning and Behavior* 4:53–57.

Newberry, B. H. 1971. Response variability and the partial reinforcement effect *Journal of Experimental Psychology* 89(1):137–141.

Pryor, K. W., Haag, R., and O'Reilly, J. 1969. The creative porpoise: Training for novel behavior. *Journal of the Experimental Analysis of Behavior* 12 653–661.

Rotter, J. B. 1966. Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs* 80(1):609.

Rotter, J. B., Liverant, S., and Crowne, D. P. 1961. The growth and extinction of expectancies in chance controlled and skilled tasks. *Journal of Psychology* 52:161–177.

Scandura, J. M. 1970. Role of rules in behavior: Toward an operational definition of what (rule) is learned. *Psychological Review* 77:516–633.

Schrier, A. M. 1974. Transfer between the repeated reversal and learning set tasks: A re-examination. *Journal of Comparative and Physiological Psychology* 87:100–110.

Skinner, B. F. 1969. *Contingencies of reinforcement.* New York: Appleton-Century-Crofts.

Surridge, C. T., and Amsel, A. 1965. Performance under a single alternation schedule of reinforcement at 24-hour intertrial intervals. *Psychonomic Science* 3:131–132.

Weiner, B., Frieze, I., Kukla, A., Reed, L., Rest, S., and Rosenbaum, R. M. 1972. Perceiving the causes of success and failure. In *Attribution: Perceiving the causes of behavior,* ed. E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, and B. Weiner. Morristown, N.J.: General Learning Press

Weiner, B., Heckhausen, H., Meyer, W., and Cook, R. E. 1972. Causal ascriptions and achievement behavior: A conceptual analysis of effort and reanalysis of locus of control. *Journal of Personality and Social Psychology* 21:239–248.

Wong, P. T. P. 1975. The concept of higher order operant: A preliminary analysis. *Bulletin of the Psychonomic Society* 5:43–44.

Wong, P. T. P. 1977a. A behavioral field approach to instrumental learning in the rat: I. Partial reinforcement effects and sex differences. *Animal Learning and Behavior* 5:5–13.

Wong, P. T. P. 1977b. Durable partial reinforcement effects and social dominance in the rat. *Learning and Motivation* 8:275–283.

Wong, P. T. P. In press. A behavioral field approach to instrumental learning the rat: II. Training ...

AMERICAN JOURNAL OF PSYCHOLOGY
December 1977, Vol. 90, No. 4, pp. 675–687

Competition between memory and perceptual tasks involving physically similar stimuli

Seth N. Greenberg
Ohio University — Lancaster

Performance on a perceptual recognition task and a concurrently performed short-term memory (STM) task was studied. The stimuli used in the STM task were either visually similar, acoustically similar, or dissimilar to the stimuli in the perceptual task. The results demonstrated that any type of concurrent list memorization has a detrimental effect upon perceptual performance, but this effect is magnified if the stimuli used in the 2 tasks are visually similar. Acoustic similarity had no effect. The results document the fact that distinct tasks interfere, when the tasks compete for a limited total processing capacity, or when the tasks demand the same visual coding program.

Competition between the distinct processes of perception and storage has been explored by Doost and Turvey (1971), Shulman and Greenberg (1971), Shulman, Greenberg and Martin (1971), and Scarborough (1972). The basic procedure in all these experiments was to present a memory list, followed by a brief visual display, and then to have the subject record the perceptual response or the memory response, or the perceptual response followed by the memory response. Shulman and Greenberg and Shulman et al. used visual displays that contained either a single item or two comparison items. Use of one- or two-item perceptual displays, in combination with an immediate response procedure, made the storage process relatively unimportant in the performance of the perceptual task. Doost and Turvey, and Scarborough, used larger perceptual displays, and so competition for storage space could explain, in part, any decrement in perceptual performance resulting from the competing short-term memory (STM) task. Doost and Turvey failed to find interference between the two

ory list is at least 5 items long. As well as Shulman and Greenberg, Shulman et al. and Scarborough discovered that the performance of an STM task interferes with the performance of a perceptual task, and at least in the case of Shulman and Greenberg, and Shulman et al., the interference could not be attributed to direct competition for a common processing function, i.e., storage.

Interference between distinct processes is consistent with a theoretical viewpoint set forth by Moray (1967). Moray argued that while processing capacity can be used wherever it is needed, the system's available capacity has an upper limit. When one task's demand on the system for a particular function is great, the capacity available for the performance of a completely different function, demanded by a second task, may be severely limited. Thus, one boundary condition set for the simultaneous performance of distinct tasks is that they cannot be performed independently when together they demand more processing capacity than the total system has available.

A second boundary condition, not explored in previous research, is the effect of stimulus similarity on the performance of competing distinct tasks. The present study explored the possibility that additional competition could be generated between a memory and a perception task if both tasks employed physically similar stimuli. It might be proposed that if the information stored for memory task performance demands a type of coding program that is also necessary for identification of the perceptual item, the magnitude of interference between the tasks would be increased above the level already reached because of the competition for total system capacity.

The memory and perceptual material could have visual and/or acoustic properties in common. Sperling (1963, 1967) suggested that following a brief visual presentation the perceptual stimulus is retained by coding the display as implicit speech. Therefore, it might be predicted that information in short-term storage (STS) that is acoustically similar to the perceptual information could prevent the necessary acoustic coding of the perceptual item because of its previous engagement of the coding program, and hence cause a greater decrement in perceptual processing than an acoustically different list.

There was no place in Sperling's model for visual features to be stored, but recent research has provided strong evidence favoring a STS for visual information. Hall, Swane, and Jenkins (1973) and Murphy (1973) have shown that information stored in a STM task is differentially affected by the performance of "visual" and "acoustic" interpolated tasks. Similarly, Pellegrino, Siegel, and Dhawan (1975, 1976) demonstrated

with a Brown-Peterson short-term retention task that picture triads were better retained than word triads when subject performed the counting backwards interpolated task. These authors argued that such evidence favors a dual encoding of figures (cf. Pavio, 1971) with visual features and acoustic features being deposited in separate storage systems. Murdock and Walker (1969) and Nilsson (1973) have found that subjects are likely to cluster information according to presentation modality. Shiffrin and Atkinson (1969) have argued that such evidence would support the notion of separate short-term acoustic and visual storage systems. If visual information can also be coded and stored in a way that maintains visual properties, clearly increased interference attributable to visual similarity between the two tasks is also a possibility.

The basic design of the experiment included 2 sets of pronounceable memory items (words); one set was acoustically confusable with half of the perceptual stimuli while the other set was not confusable with any perceptual stimuli. Two sets of visual memory items (random forms) were also used; one set was visually confusable with the perceptual stimuli, while the second set was not. Efforts were made to assure that the processing demands made by the 2 sets of random forms were equal, and the demands made by the 2 types of word sets were equal.

In addition, 2 delay conditions were used. The subject was either required to respond immediately to the perceptual task or to hold the perceptual information together with the memory information for 7 sec. The delay condition was included to check the possibility that any similarity effect that developed was due mainly to interference between the stored memory items and the perceptual item while it was maintained in storage, rather than to competition for a particular coding program. If a perceptual decrement resulting from similarity is attributable to interference between information being stored the magnitude of the effect should be greatly increased in the delay condition. If, on the other hand, the decrement occurs because the proper coding program is not available at the time of perceptual coding, delay of report would have relatively little impact on perceptual performance.

METHOD

Subjects

Twelve undergraduate students served as subjects in order to earn extra credit for a psychology course.

Apparatus

The basic projection equipment included a Kodak RA950 projector located behind a Gerbrand G1166 electronic shutter and a Kodak 800 Carousel projec-

tor located behind a blade shutter. Both projectors were controlled by a Psionix 1248A timer.

Perceptual stimuli

Two sets of perceptual stimuli were employed: random forms (nonpronounceable) and pronounceable forms. The perceptual group consisting of 3 random forms was derived from a square pattern (Figure 1, left). The pronounceable group consisted of 3 acoustically similar items — the letters A and H, and the number 8 — all drawn in block form, so as to make them visually similar as well. Note that the items in the pronounceable group (A, H, and 8) were also derivatives of the basic square pattern. Thus, the perceptual stimuli consisted of 2 different groups of square items: one group with nonpronounceable items, the second group with items that all begin with the sound [a] as in *ate*. Examples of perceptual stimuli are given in Figure 2.

Memory stimuli

Four types of memory stimuli were used: square figures, curved figures, "a" words, and "e" words. Fifteen square figures were derived from the same square pattern used to produce the 6 perceptual stimuli, while 15 curved figures were derived from a curved pattern (see examples in Figure 3). The square and curved figures plus 10 miscellaneous figures were rated for similarity to the most complex random figure being used as a perceptual stimulus. Twenty-five students, none of whom participated in the major experiment, provided ratings. The 10 square figures that were rated most similar and the 10 curved figures rated least similar formed the pools of square and curved memory stimuli used in the experiment. On a scale of 5, no square figure of the 10 chosen had a mean rating below 4, with the average rating for the group of 10 figures 4.5, and no chosen curved figure was rated higher than a 2.2, with the average curved figure rating 1.8. In addition, the degree of maximum overlap between each of the chosen memory figures and the perceptual items was determined. The average overlap for the 10 square figures when compared to each of the perceptual items exceeded 75% in each case, while average overlap for the 10 curved figures in no case exceeded 20%. Thus, it could be stated that the square memory figures were

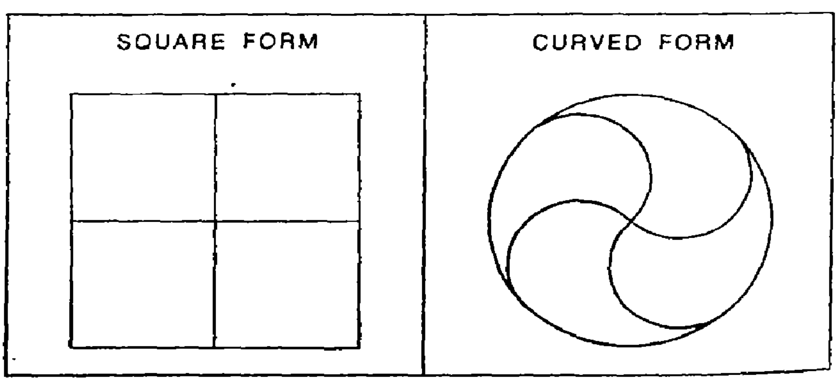


Figure 1. Basic patterns for perceptual stimuli

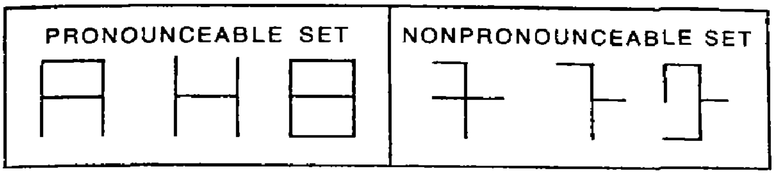


Figure 2. Examples of perceptual stimuli

visually confusable with both sets of perceptual items, while the curved figures were not.

The "a" and "e" words met 4 requirements: (1) words had to be less than 7 letters long and contain no more than 2 syllables, (2) each word had to begin with a long "a" sound as in *ate* or a long "e" sound as in *east,* (3) words could not be nouns, (4) the words had to be low-image. To test for imagery, 50 students, none of whom participated in the major experiment, were given the imagery instructions used by Pavio, Yuille, and Madigan (1968) and asked to rate 14 "a" words, 13 "e" words, and 13 additional words that varied considerably in imagery, according to the Pavio, Yuille, and Madigan norms. The average imagery rating for the 10 "a" and 10 "e" words selected for the experiment was 2.5 and 2.2, respectively. The "a" words were acoustically similar to the pronounceable perceptual items, while the "e" words were not.

Design

Each subject participated in 6 one-hour sessions on consecutive weekdays. The first 2 sessions were given to the determination of 90% duration thresholds using a short-cut psychophysical procedure described by Reicher (1968), and to practice on the memory task alone and both the memory and perceptual tasks together. The next 4 sessions each consisted of 48 experimental and 12 control trials.

Each subject served in every condition of the experiment. There were 16 experimental conditions representing the factorial combination of response delay (0 sec and 7 sec) with perceptual type (random forms and pronounceable items) with memory list type ("a" words, "e" words, square figures and curved figures). It was not necessary to factorically combine all conditions for control conditions since the purpose of the control condition was to separate the perceptual and memory tasks. Thus the control conditions called for the factorial combination of delay with memory type and delay with perceptual type, but not memory type with perceptual type.

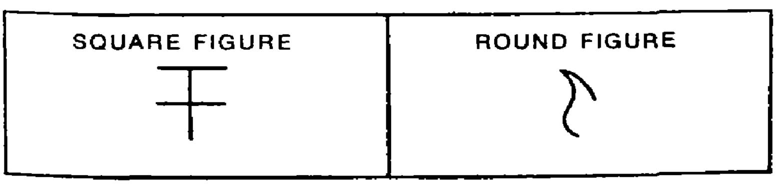


Figure 3. Examples of memory stimuli figures

Each delay condition was presented for 2 consecutive sessions. Each perceptual type was used either the first or second half of a session for all 4 sessions. The presentation orders that resulted from the factorial combination of delay with perceptual type were counterbalanced across subjects.

Each half session began and concluded with 3 control trials; sandwiched between were 24 experimental trials. There was a 5-min break between session halves. Within each 24-experimental-trial block, each type of memory list was employed twice in combination with each of the 3 items composing the pronounceable perceptual group, and twice with each of the 3 items making up the random perceptual group. On one of the occasions in which a particular memory type was combined with one of the perceptual items a positive memory recognition response was required, while on the other occasion a negative response was appropriate. Each memory list was used equally often during the 12 control trials. The order for presenting the 4 types of memory lists was randomized over a session.

Memory lists were 5 items long. The assignment of items to the memory lists, for all memory types, assured that over 2 consecutive sessions each item appeared with approximately equal frequency in each serial position. No item appeared more than once in a list. Each 5-item list was used twice, once each odd numbered session or once each even numbered session.

The 6 perceptual stimuli were exhibited equally often under each factorial combination of memory type with delay for the experimental trials, and equally often with each delay condition during the control trials. With the above restrictions met, the order of presentation of the perceptual stimuli was randomized.

Procedure

The experimental trial began when the subject indicated by saying "ready" that he was fixated on a small rectangular area on the right side of the projection screen. Immediately, the 5 items composing the memory list for that trial were simultaneously displayed, in a vertical column, 3 in. to the left of the fixation area, for a period of 12 sec. The memory list was clearly visible to the subject even while the subject focused on the fixation area. All 5 items composing the memory list were the same type of item, e.g., all were "a" words. One second after the memory list ended, the perceptual display containing a single random form or pronounceable item was presented in the fixation area. During the 1-sec period between the end of the memory display and the beginning of the perceptual display the screen was dark. Following termination of the perceptual stimulus a mask composed of broken *X*s and *O*s was displayed for 3 sec in the fixation area. The fixation area subtended a visual angle of approximately 2.8°. Instructions for the experimental trial emphasized to the subject that he or she "keep the memory items in mind, while looking at the perceptual item." Either immediately after the perceptual display ended or 7 sec later, the subject was signaled to respond by a 750-Hz tone presented for 1 sec. The response to a trial was made on two sides of a page in a 3-by-5-in. booklet. On one side of the page was the appropriate group of perceptual stimuli, either 3 random figures or 3 pronounceable figures, while the other side of the page contained a memory list that either matched the just-presented list perfectly or deviated from the presented list because one of the presented items was replaced by an item of a similar type that had not been presented on that trial. On experimental trials the subject was

required first to check the perceptual item he saw displayed and then to flip to the other side of the page and check either a box at the bottom of the page, indicating that all items matched the presented memory list, or check the substituted item. Response time was generally less than 20 sec. The subject always knew in advance from which of the 2 perceptual groups the perceptual item would be chosen, but did not know in advance which type of memory list was about to be presented. In addition, the subject knew in advance whether the trial required an immediate or delayed response. In the case of delayed responses the subject was asked to keep the top of the response page covered with a hand until the tone sounded, thus discouraging the subject from using positional cues to aid in the delayed response. The positions of the perceptual stimuli on each page were randomly assigned.

On the control trials, performance was required on both the perceptual and memory tasks, but performance of the memory task (including the response) preceded the presentation of the perceptual stimulus. As with the experimental trial, the control trial began after the subject indicated fixation by saying "ready." The sequence and timing of events composing the control trial followed the course already described for the experimental trial, except that in the control trial a gray slide was substituted for the perceptual item. Either immediately or 7 sec after the masking slide terminated, the subject responded to the memory stimulus by flipping the appropriate page and checking the correct box or substitute stimulus. Following that response, the control trial continued. The preceding sequence of events was repeated, but this time the gray slide was substituted for the memory list, and the perceptual stimulus was shown. The subject responded to the perceptual stimulus on the opposite side of the page just used for the memory response, and the control trial was completed. As was true with the experimental trial the subject knew in advance the type of delay being used, and the type of perceptual item, but not the type of memory list.

RESULTS

Perceptual processing was significantly affected by the competing memory task, $F(4, 44) = 18.10, p < .01$. A Newman-Keuls analysis (Winer, 1962, pp. 85–89) of the main effect of type of memory list indicated that when the perceptual task was performed in combination with the memorization of any of the 4 types of lists, significantly more perceptual errors were made than when the perceptual task was performed alone (control condition). All of the significant Newman-Keuls comparisons were significant at the .01 level.

Two additional analyses of variance were performed. The first analysis made a direct comparison of the relative impact that square-figure and curved-figure memorization had on the perception of both pronounceable and random forms. This analysis revealed that square-figure memorization led to significantly more perceptual errors, with either perceptual set, than curved-figure memorization, $F(1, 11) = 28.73, p < .01$. There were no significant interactions.

The second analysis compared the relative effects of "a" word and "e" word memorization on perception of pronounceable items. Despite the similarity in sound between "a" words and the items included in the pronounceable set, there was no indication that "a" words produced more perceptual errors than "e" words. Again no significant interactions were produced.

All 3 analyses failed to show any main effect or interaction involving delay. The major analysis and first supplementary analysis did, however yield a significant main effects of perceptual type $F(1, 11) = 7.40, p < .05$, and $F(1, 11) = 7.75, p < .05$, respectively. The main effect of perceptual type reflects a lack of precision in establishing a 90% duration threshold. This is clearly indicated in Table 1, which summarizes the error data for the perceptual task. Note that a large difference between the random and pronounceable perceptual types existed even in the control condition, when an immediate response was required (a condition very similar to the threshold determination situation).

A conservative criterion was used to measure recognition performance. A trial requiring a negative response was considered correct only if the subject correctly identified the substitute item; if the subject indicated the list was incorrect but failed to identify the substitute item the response was wrong. Table 2 summarizes the error data for recognition performance in the memory task as a function of memory list type and type of interpolated task, i.e., perceptual task with random forms, perceptual task with pronounceable items, and no perceptual task (control condition)

Type of interpolation was significant, $F(2, 22) = 7.69, p < .01$, indicating that memory performance was superior when the task was performed alone rather than in combination with a perceptual task However, a significant interaction of type of interpolation with type of memory

Table 1. Perceptual error rates in percentages as a function of memory-list type, delay, and perceptual-stimulus type

	Type of memory list				
Delay	Square form	Curved form	"a" word	"e" word	No list control
Pronounceable perceptual type					
0 sec	23	15	19	17	7
7 sec	28	21	17	25	16
Random perceptual type					
0 sec	35	28	23	28	14
7 sec	39	26	24	24	18

Table 2. Memory error rates in percentages as a function of memory-list type, delay, and type of interpolation

Delay	Type of memory list: Square form	Curved form	"a" word	"e" word
Pronounceable stimuli				
0 sec	31	29	10	13
7 sec	29	30	13	15
Random stimuli				
0 sec	34	27	9	6
7 sec	34	29	7	6
No perceptual stimuli (control)				
0 sec	18	15	5	8
7 sec	23	21	3	11

list, $F(6, 66) = 3.66, p < .01$, and an inspection of Table 2 suggested that the inclusion of a perceptual task affected memory performance only when square or curved form lists were used; not when word lists were used. A comparison of the error rates by the Newman-Keuls procedure confirmed this interpretation. The difference between the respective control condition error rates and their experimental condition counterparts was significant ($p < .01$) when form lists were involved, but not when word lists were used.

To test for similarity effects, 2 supplementary analyses were again performed. The first analysis compared square-figure and curved-figure recognition performance when the memory task was performed with perception of either pronounceable or random forms. The second analysis compared "a" and "e" word recognition when performed in combination with perception of pronounceable items. The analyses also tested for effects of delay. Neither analysis revealed any effects involving similarity or delay.

Finally, the major analysis of variance demonstrated an effect of type of memory list, $F(3, 33) = 24.24, p < .01$, reflecting the fact that regardless of whether an interpolated task was performed or not, figure memorization was more difficult than word memorization.

DISCUSSION

In conformity with previous research, the present study found storage of information in an STM task to have a detrimental effect upon perceptual processing in a detection task. In addition, it was determined that

the magnitude of the effect was related to the degree of visual similarity, but not acoustic similarity, between the stored information and the perceptual information. Whereas the decrement in perceptual processing caused by the storage of any of the 4 types of memory lists suggests that the 2 distinct processes, perceptual encoding and rehearsal of stored information, cannot proceed independently once some upper limit on total processing capacity is surpassed (Shulman and Greenberg, 1971), the additional decrement associated with the storage of the visually similar list argues for specific competition between the 2 tasks for a particular visual coding program. Alternatively, it might be argued that the visually similar material produced more of a decrement than the visually different list because it demanded more total processing capacity. The alternative explanation is challenged by 3 things. First, construction of the square and curved figures was done in such a way as to make both groups of stimuli equally complex. Second, memory performance with both lists in the control condition was equivalent, indicating the lists were equally difficult. Finally, a pilot study showed that the 2 types of lists produced a significant, but equivalent, decrement in a neutral perceptual task, i.e., where physical similarity was not a factor (recognition of 6, 7, or 8 dots in a brief visual display), $F(2, 6) = 6.32, p < .05$.

The difference in the magnitude of competition resulting from square- and curved-figure memorization also indicates a new type of evidence favoring the existence of a short-term storage system for visual properties. If such a storage system did not exist, the general visual characteristics, i.e., squareness and curvedness, could not have resulted in different effects upon perceptual performance.

As in Scarborough (1972), when acoustically codable memory lists were involved, competition between the tasks was found to affect only perceptual performance; however, when visually coded STM lists were used, the present study found that performance on the memory task also suffered. Unlike percepual performance, memory performance was not differentially affected by visual similarity. That is, square-form memorization was not influenced more by the perceptual task than curved-form memorization. The significant, but nondifferentiated, effect upon form memorization indicates that in this experiment the visually coded STM lists probably required more processing capacity for rehearsal than the acoustically codable lists. It also indicates that STM performance was not influenced by the competition for a particular visual coding program.

The fact that STM performance was not affected by competition for the coding program, while perceptual performance was, and the finding that STM performance did not suffer when word lists were used, while

perceptual performance did, raises a question as to why the perceptual task took the brunt of the competition. Norman and Bobrow (1975) explained that "asymmetry of interference between two tasks is likely to depend upon task instructions and subject strategy — upon which of the competing tasks receives first priority." The instructions used in the present experiment did emphasize memory performance, so, according to Norman and Bobrow, the asymmetric results were to be expected.

While involvement with a STM task had a significant effect upon perceptual performance, that effect did not significantly increase with a 7 sec delay, i.e., there was no condition-by-delay interaction, $F < 1$. Evidently the major impact upon perceptual processing of a single item occurs before or during perceptual encoding, and not during storage. Aaronson (1967) had argued that it is before information is completely encoded, when it is in an unstable state, that the occurrence of other events has its greatest impact. The present results demonstrated that while the visually presented perceptual information is in this unstable state, waiting to be coded, ongoing activity that denies the perceptual information the proper visual coding program is particularly devastating to the perceptual process. This is true even when the perceptual information can be named. Posner, Boies, Eichelman, and Taylor (1969) provided evidence that indicated that visual properties of visually presented material are likely to be processed before acoustic (name) properties. The present study suggests that the initial coding of visual properties is essential to the subsequent coding of acoustic properties. In the present study, storage of nonpronounceable visually similar forms had as great an effect upon pronounceable as upon nonpronounceable perceptual information (the conditions-by-perceptual-stimulus-type interaction produced an $F < 1$). Presumably, if the use of an acoustic code is independent of visual coding, then the identification of A, H, and 8 could proceed uninhibited by the visual similarity of the stored information. In contrast, random figures that are solely dependent upon visual coding should be affected to a much greater extent than the pronounceable items by the visual similarity of STM information. This was not the case.

To summarize, though perceptual and memory tasks are considered distinct, they cannot be performed independently when together they demand more processing capacity than the system has available, or when the stimuli used in the 2 tasks have visual properties in common. Furthermore, it is the coding of the visual properties that enables a visually presented perceptual item to be identified; acoustic coding is relatively unimportant.

Notes

This research was supported in part by funds obtained from the Ohio University Research Council. The author wishes to acknowledge the assistance of Burt Haller in preparing and executing this experiment, and Pat Milbaugh for preparing the manuscript. The author also wishes to thank Harvey Shulman and Frank Bellezza for their helpful comments on an earlier version of this paper.

This paper was presented in part at the Midwestern Psychological Association Convention, May 1976. Requests for reprints should be sent to Seth N. Greenberg, Department of Psychology, Ohio University — Lancaster, 1570 Granville Pike, Lancaster, OH 43130. Received for publication November 18, 1976; revision, February 14, 1977.

References

Aaronson, D. 1967. Temporal factors in perception and short-term memory. *Psychological Bulletin* 67:130–144.

Conrad, R. 1964. Accoustic confusions in immediate memory. *British Journal of Psychology* 55:75–84.

Doost, R., and Turvey, M. T. 1971. Iconic memory and central processing capacity. *Perception and Psychophysics* 9:269–274.

Hall, R., Suane, D., and Jenkins, R. A. 1973. Memory for physical and semantic features of visual material in a shadowing task. *Journal of Experimental Psychology* 100:426–428.

Moray, N. 1967. Where is capacity limited? A survey and model. *Acta Psychologica* (Amsterdam) 27:84–92.

Murdock, B. B., Jr., and Walker, K. D. 1969. Modality effects in free recall *Journal of Verbal Learning and Verbal Behavior* 8:665–676.

Murphy, R. J. 1973. Recognition memory for sequentially presented pictorial and verbal spatial information. *Journal of Experimental Psychology* 100: 327–334.

Nilsson, L. 1973. Organization by modality in short-term memory. *Journal of Experimental Psychology* 100:246–253.

Norman, D. A., and Bobrow, D. G. 1975. On data-limited and resource limited processes. *Cognitive Psychology* 7:44–64.

Pavio, A. 1971. *Imagery and verbal processes.* New York: Holt, Rinehart, and Winston.

Pavio, A., Yuille, J. C., and Madigan, S. 1968. Concreteness, imagery and meaningfulness values for 925 nouns. *Journal of Experimental Psychology* 76 (1, Pt. 2).

Pellegrino, J. W., Siegel, A. W., and Dhawan, M. 1975. Short-term retention of pictures and words: Evidence for dual coding systems. *Journal of Experimental Psychology: Human Learning and Memory* 104:95–102.

Pellegrino, J. W., Siegel, A. W., and Dhawan, M. 1976. Short-term retention of pictures and words as a function of type of distraction and length of delay interval. *Memory and Cognition* 4:11–15.

Posner, M. I., Boies, S. J., Eichelman, W. H., and Taylor, R. L. 1969. Retention of visual and name codes of single letters. *Journal of Experimental Psychology Monograph* 79 (No. 1, Pt. 2).

Reicher, G. M. 1968. *Perceptual recognition as a function of meaningfulness of stimulus material* (Tech. Rep. No. 7). Ann Arbor: University of Michigan Human Performance Center.

Scarborough, D. L. 1972. Memory for brief visual displays of symbols. *Cognitive Psychology* 3:408–429.

Shiffrin, R. M., and Atkinson, R. C. 1969. Storage and retrieval processes in long-term memory. *Psychological Review* 76:179–193.

Shulman, H. G., and Greenberg, S. N. 1971. Perceptual deficit due to division of attention between memory and perceptions. *Journal of Experimental Psychology* 88:171–176.

Shulman, H. G., Greenberg, S. N., and Martin, J. 1971. Intertask delay as a parameter of perceptual deficit in divided attention. *Journal of Experimental Psychology* 88:439–440.

Sperling, G. 1963. A model for visual memory tasks. *Human Factors* 5:19–31.

Sperling, G. 1967. Successive approximations to a model for short-term memory. *Acta Psychologica* 27:285–292.

Winer, B. J. 1962. *Statistical principles in experimental design*. New York: McGraw-Hill.

continued from page 634

Schilder, Paul. *On Psychoses.* New York: International Universities Press, 1977. Pp. 594. $22.50.

Schmidt, Jerry A. *Help Yourself: A Guide to Self-Change.* Champaign, Ill.: Research Press, 1976. Pp. 128. $3.95.

Schmitt, Cecelia. *Rapport and Success.* Ardmore, Pa.: Dorrance, 1976. Pp. 213. $4.95.

Schneidman, Edwin S. *Death: Current Perspectives.* Palo Alto, Calif.: Mayfield, 1976. Pp. 547. Paperback. $7.95.

Schneidman, Edwin S. *Suicidology: Contemporary Developments.* New York: Grune & Stratton, 1976. Pp. 571. $27.50.

Schoenberg, Bernard, et al. *Bereavement: Its Psychological Aspects.* New York: Columbia University Press, 1976. Pp. 375. $15.00.

Schutz, Will, and Turner, Evelyn. *Evy: An Odyssey into Bodymind.* New York: Harper & Row, 1976. Pp. 135. $6.95.

Scitovsky, Tibor. *The Joyless Economy: An Inquiry into Human Satisfaction and Consumer Dissatisfaction.* New York: Oxford University Press, 1976. Pp. 310. $11.95.

Scott, John Paul. *Aggression.* Rev. ed. Chicago: University of Chicago Press, 1976. Pp. 233. Paperback. $4.95.

Scott, Rosemary. *The Female Consumer.* New York: Halsted, 1976. Pp. 352. $24.95.

Sears, Robert R.; Maccoby, Eleanor E.; and Levin, Harry. *Patterns of Child Rearing.* Stanford, Calif.: Stanford University Press, 1976. Pp. 549. $17.50.

Sebald, Hans. *Momism.* Chicago: Nelson-Hall, 1976. Pp. 318. $8.95.

Semeonoff, Boris. *Projective Techniques.* New York: Wiley, 1976. Pp. 336. $10.95.

Shapiro, Stephen, and Ryglewicz, Hilary. *Feeling Safe.* Englewood Cliffs, N.J.: Prentice-Hall, 1976. Pp. 148. $8.95.

Sheehan, Peter W., and Perry, Campbell W. *Methodologies of Hypnosis.* New York: Halsted, 1976. Pp. 329. $18.00.

Simkin, James S. *Gestalt Therapy: Mini-lectures.* Millbrae, Calif.: Celestial Arts, 1976. Pp. 124. Paperback. $3.95.

Simon, Sidney B., and Olds, Sally Wendkos. *Helping Your Child Learn Right from Wrong.* New York: Simon & Schuster, 1976. Pp. 223. $7.95.

Siu, R. G. H. *Ch'i: A Neo-Taoist Approach to Life.* Cambridge, Mass.: MIT Press, 1976. Pp. 351. Paperback. $4.95.

Skirbekk, Sigurd. *Psychoanalysis and Self-understanding.* Atlantic Highlands, N.J.: Humanities Press, 1976. Pp. 151. $8.00.

Slater, P., ed. *Explorations of Intrapersonal Space.* New York: Wiley, 1976. Pp. 258. $19.95.

Smith, Edward W. L., ed. *The Growing Edge of Gestalt Therapy.* New York: Brunner-Mazel, 1976. Pp. 239. $12.95.

Smith, Judith M., and Smith, Donald E. P. *Child Management.* Champaign, Ill.: Research Press, 1976. Pp. 144. Paperback. $3.95.

Smith, W. Lynn, and Kling, Arthur, eds. *Issues in Brain/Behavior Control.* New York: Halsted, 1976. Pp. 157. $12.50.

Snijders, J. Th., and Snijders-Oomen, N. *The Snijders-Oomen Non-verbal Intelligence Scale.* Groningen, Holland: H. D. Tjeenk Willink bv, 1976. Pp. 60. Paperback.

AMERICAN JOURNAL OF PSYCHOLOGY
December 1977, Vol. 90, No. 4, pp. 689–698

The relative effectiveness of rules 1 and 2 in verbal discrimination learning

Peter M. Lisiecki and Terry M. Libkuman
Central Michigan University

The prediction, derived from frequency theory, that performance on a verbal discrimination task would be the same for Rules 1 and 2 was tested. Appropriateness of the rules was achieved by varying the number of times the right (R) and wrong (W) items were pronounced during the feedback interval. The sequence of pronunciation (R-W or W-R) was also manipulated in order to determine whether rehearsal depends on the order of pronunciation. After the task was completed, each subject was given a modified free-recall task to investigate possible relationships between the various treatment conditions and incidental learning. Analyses of the data revealed a significant differential pronunciation effect ($p < .05$) indicating that Rule 2 responding retarded acquisition rate. Superior performance on the modified free-recall task was associated with a W-R, relative to an R-W, sequence of pronunciation during feedback. Discussion of the results indicated possible explanations inconsistent with frequency theory predictions.

The frequency theory of verbal discrimination (VD) learning (Ekstrand, Wallace, and Underwood, 1966) proposes that the cue for discrimination is the subjective difference in frequency of occurrence between the right (R) and wrong (W) alternatives of each stimulus pair. The differential accrual of frequency units to the R and W items is assumed to be the result of 4 components operating during the VD task: representational responses (RR), pronunciation responses (PR), rehearsal-of-the-correct responses (RCR), and implicit associative responses (IAR).

Frequency theory assumes that the subject uses one of two "rules" to assist him in making a discrimination: Rule 1 (select the more frequent item of a pair) or Rule 2 (select the less frequent item of a pair). Since in the typical learning situation the rate of accrual of frequency favors the R item (via PRs and RCRs), the subject must adopt a Rule 1 mode of responding to master the VD task. According to Ekstrand et al. (1966), however, "There is no reason to believe that when frequency difference is

reversed (i.e., the less frequent member of each pair is correct), subjects cannot operate using Rule 2" (p. 569).

Several experiments in which transfer paradigms were used (Lovelace, 1966; Underwood and Freund, 1968, Experiment I; Underwood, Jesse, and Ekstrand, 1964) have provided information pertinent to the use of Rule 2. These authors obtained results indicating initial positive transfer followed by a decrement in performance and/or subsequent slow improvement across trials when the situation required a Rule 2 mode of responding. The results are entirely supportive of predictions derived from frequency theory. On the initial trials of the transfer list, always selecting the less frequent item of a pair resulted in the correct response. However, as learning progressed, frequency accumulated to the R item faster than to the W item. This reduced the frequency difference between the R and W alternatives, thus impeding the rate of VD acquisition.

In a recent study, Smith and Jensen (1971) allowed 3 groups of subjects an equal degree of practice on either correct, incorrect, or both correct and incorrect items prior to VD learning. This study is relevant here because Smith and Jensen hypothesized that the correct group would identify Rule 1 more quickly than the incorrect group would identify Rule 2. Consequently, this was expected to facilitate performance of the correct group relative to the incorrect group on the first test trial. While the results confirmed the authors' prediction it is our contention that such a hypothesis is not derivable from frequency theory. The theory can predict the direction of performance only on the basis of the frequency differential between the alternatives of a pair. That is, on the initial test trial the correct and incorrect groups were operating under equivalent frequency differentials; therefore, frequency theory predicts equivalent performance from these groups.

Further evidence that subjects can and do use Rule 2 was reported by Newby and Young (1972), who manipulated frequency within a single VD list. They also found learning of Rule 2 groups to be retarded subsequent to the first test trial (where performance was comparable to that of subjects using Rule 1). Furthermore, Newby and Young theorized that the performance of the Rule 2 groups would have paralleled that of the Rule 1 groups had the former been able to continue using Rule 2 throughout the VD task (i.e., had not the relatively greater accumulation of frequency to the R item forced subjects to switch to Rule 1 responding).

While the results are consistent with predictions derived from frequency theory, they do not appear to be a valid test of the effectiveness with which a subject can apply Rule 2 to a VD list. That is, in none of these studies was the use of Rule 2 the focus of investigation. Consequently, the implication that either rule can be applied with equal facility was

based on the generally consistent finding of no difference in performance during the initial trials only.

In a recent study, Newman, Suggs, and Averitt (1974, Experiment I) investigated the use of Rules 1 and 2 over 4 trials of learning. Appropriateness of Rule 1 was manipulated by having the subject indicate on test trials the *same* item of a pair designated by the experimenter on study trials. Appropriateness of Rule 2 was achieved by having the subject indicate on test trials the item of a pair *different* from that designated by the experimenter on study trials. The authors used 4 treatment conditions: RR, WW, RW, and WR. The first letter refers to the item designated by the experimenter on study trials. The second letter refers to the item that the subject was instructed to indicate on test trials. The results indicated inferior performance in the RW and WR groups (Rule 2) when compared to the RR and WW groups (Rule 1). The differences were apparent on the first trial, with both groups displaying a comparable rate of improvement over the 4 trials of learning. Thus, it appears that the aforementioned "rule switching" behavior cannot account for the relatively poorer performance of the Rule 2 groups. Newman et al. suggested, however, the possibility that Rule 1 was actually the more appropriate in all 4 groups. This would result if individuals covertly rehearsed the items they were instructed to indicate on test trials rather than the items designated by the experimenter on study trials. The better performance of the RR and WW groups relative to the RW and WR groups would be due to a greater difference in frequency between the R and W items in the former groups. Moreover, even if covert rehearsal did not contaminate the results as such, the failure of Newman et al. to ensure that Rules 1 and 2 were equally appropriate precluded a valid evaluation of their application to a VD list.

The theoretical implications of these results are that the relative merits of the use of Rules 1 and 2 can be properly assessed only if a constant R:W ratio is maintained throughout the VD task. This is possible, to some extent, by equalizing the rate with which frequency is added to the R and W items (e.g., controlling the emission of PRs and RCRs). Moreover, according to a strict frequency theory interpretation, a R:W ratio of 2:1 or 1:2 should have the same effect on VD acquisition.

The purpose of the present paper was to determine the relative effectiveness of the use of Rules 1 and 2 in VD learning. It was hypothesized that no difference would exist between groups where the use of the 2 modes of responding (i.e., Rule 1 and Rule 2) were equally appropriate.

The design was a 2 × 4 factorial. The first factor was sequence of pronunciation during the feedback interval (R-W or W-R). This variable was included to control for the possibility that subjects instructed to pro-

nounce the W item and then the R item (W-R) might be more likely to covertly rehearse the R item than subjects given the reverse order (R-W). The second factor was differential pronunciation (viz., overt pronunciation of the VD items during the feedback interval). More specifically, Rules 1 and 2 were manipulated by having subjects respond to each item of the VD pair a specific number of times. Subjects were assigned to one of 4 groups designated as follows: High Ratio Right (HRR); High Ratio Wrong (HRW); Low Ratio Right (LRR); and Low Ratio Wrong (LRW). Theoretically, and as a result of the frequencies accrued during the anticipation and feedback intervals, the R:W ratios associated with these 4 groups were approximately 8:3, 3:8, 5:3, and 3:5 respectively. Manipulation of this variable resulted in Rule 1 being as appropriate to the HRR and LRR groups as Rule 2 was to the HRW and LRW groups respectively.

Valid conclusions regarding the relative effectiveness of Rules 1 and 2 are limited to the results based on comparisons within the high ratio groups (HRR vs. HRW) and within the low ratio groups (LRR vs. LRW) since the frequency differences remain constant for these comparisons. The HRR vs. LRW and the HRW vs. LRR comparisons are not appropriate since there is a greater difference in frequency between the R and W items in the 2 high ratio groups. Therefore, according to frequency theory it was predicted that no differences would exist between the 2 high ratio groups or between the 2 low ratio groups. Furthermore since there was a greater frequency differential between the R and W items in the HRR and HRW groups, performance should be facilitated in these groups when compared to the LRR and LRW groups.

Finally, in a previous study Sardello and Kausler (1967) found modified free recall (MFR) of intrapair associations to increase with overt pronunciation relative to no pronunciation of VD items in feedback. Also, an increase in recall of both R and W items was associated with the degree of practice (number of trials) on the VD task. In the present study each subject was given a MFR task upon attainment of the VD learning criterion. This task was included to ascertain whether or not there was any relationship between incidental learning and the various treatment conditions.

METHOD

Subjects

Eighty undergraduate student volunteers were selected from psychology and education courses at Central Michigan University. Subjects were randomly assigned to treatment conditions with the restriction that each of the 8 conditions consist of 6 females and 4 males. Twenty-three subjects received credit for par-

ticipation in this study; the remaining 57 received no credit. One subject failed to complete the task and was replaced by another student from the same source. All participants were naïve to verbal learning experiments.

Lists

A pool of 100 high-frequency words (1-10 per million G-count) was selected from the Thorndike and Lorge tables (1944). The words were 2-syllable nouns of 5 to 9 letters each. From this pool 50 words were randomly selected and paired to construct a 25-pair list. The selection of the R item in each pair was also determined randomly. The list was arranged in 4 random serial orders with the restriction that the R or W item occur no more than 3 times in succession on the same side of the list. Interpair and intrapair associations were eliminated via visual inspection.

Procedure

The VD items were presented via a dual-timer Marietta memory drum. A 1.5:5-sec (anticipation:feedback) rate and an 8-sec intertrial interval were used in all treatment conditions. During the anticipation interval subjects indicated by responding aloud which member of the pair they believed to be the R item. A brief interval was used to control for covert rehearsal, since Radtke, McHewitt, and Jacoby (1970) have suggested that such rehearsal may not be limited to the feedback interval.

Immediately following the anticipation interval each VD pair was exposed for 5 sec with the R item underlined. The length of the feedback interval was decided on the basis of two considerations. First, the subject was permitted sufficient time to emit the number of PRs necessary to achieve the appropriate frequency ratios. Second, the intention was to minimize the opportunity for covert rehearsal by filling the study interval as completely as possible with overt pronunciation.

In addition to the usual instructions for VD learning, the subjects assigned to the HRR group were instructed to pronounce the R item 6 times and the W item twice during the feedback interval. Those in the HRW group pronounced the R and W items 1 and 7 times respectively. Individuals assigned to the LRR group pronounced the R item 3 times and the W item twice. Finally, subjects in the LRW group pronounced the R and W items 1 and 4 times respectively.

In order to determine the approximate frequencies for the various R:W ratios it was assumed that 2 frequency units (one RR and one PR) were added to the R item, whereas only 1 unit (via the RR) was added to the W item during the anticipation interval. At this point there existed a 2:1 difference favoring the R item. Then the following R:W ratios associated with the 4 conditions were achieved by combining the PR frequencies in feedback with the frequency accrued during the anticipation interval: HRR 8:3, HRW 3:8, LRR 5:3, and LRW 3:5. Of course, these ratios would change, particularly during the earlier trials, as a function of pronouncing the W item during the anticipation interval. These changes would initially provide a slight advantage to the Rule 2 groups.

While covert rehearsal had probably been substantially reduced or eliminated in the HRR and HRW groups, such rehearsal may have occurred in the LRR and LRW groups. Consequently, the most valid test in determining the relative effectiveness of Rules 1 and 2 was considered to be the comparison between the 2 former groups.

In each treatment the sequence of pronunciation of VD items was R-W for half the subjects; for the others the sequence was W-R. For example, those in the HRR group who were instructed to pronounce the R item first (R-W) responded with the R item 6 consecutive times followed by 2 repetitions of the W item.

Subjects were informed of the criterion of 2 perfect successive trials. However, learning was discontinued if criterion was not reached by 25 trials.

Finally, subjects were given a MFR task. Each person was shown a list of stimulus items followed by a blank space for the recalled item. No time limit were set for the completion of the task. The order of MFR was counterbalanced such that half of the subjects in each group were asked to recall R items in the presence of W items first, followed by recall of W items in the presence of R items; the other half had the sequence reversed.

RESULTS

A 2 × 4 analysis of variance (sequence of pronunciation × differential pronunciation) was applied to the mean number of trials to criterion. The analysis indicated that differential pronunciation was a significant source of variation, $F(3, 72) = 9.09, p < .001$. No significant effects were found for sequence of pronunciation or the interaction. In order, the mean number of trials to criterion for the HRR, HRW, LRR, and LRW groups was: 10.90, 17.55, 11.10, and 14.60. The Newman-Keuls test for multiple comparisons yielded significant differences ($p < .05$) between Rule 1 and Rule 2 groups (HRR vs. HRW and LRR vs. LRW), but not within Rule 1 or Rule 2 groups (HRR vs. LRR and HRW vs. LRW). In general the same findings were obtained when the percentage of errors to the base of opportunity and mean number of errors to criterion were used as dependent variables. One exception, observed in both analyses, was the failure to find a significant difference between the LRR and LRW groups (the difference fell just below significance in the errors analysis). Also the errors analysis yielded a significant difference ($p < .01$) within the Rule 2 groups (HRW > LRW).

A 2 × 4 × 10 mixed-effects analysis of variance (sequence of pronunciation × differential pronunciation × trials) was applied to the mean number of errors per trial to examine the rate of VD acquisition over the first 10 trials of learning. This analysis was intended to provide a more sensitive measure of the relationship between the 4 treatment conditions. Significant main effects were found for differential pronunciation, $F(3, 72) = 12.20, p < .001$, and trials, $F(9, 648) = 265.99, p < .001$. The differential pronunciation × trials interaction was also significant, $F(27, 648) = 4.06, p < .01$. No other main effects or interactions were significant. The analysis of simple main effects revealed no significant differences between the Rule 1 groups and their Rule 2 counterparts over the first 3

trials (see Figure 1). On Trial 4, however, it is clear that Rule 2 responding has a retarding effect on VD acquisition. This is most obvious in the HRW group where there are actually more total errors on this trial than on Trial 3. This breakdown in performance is followed by a protracted period of slow improvement across the remaining trials, although it is apparent that the HRW group is having more difficulty than the LRW group.

Finally, a 2 × 4 × 2 mixed-effects analysis of variance, with sequence of pronunciation and differential pronunciation as between-group variables, and direction of recall (R-W or W-R) as a within-group variable

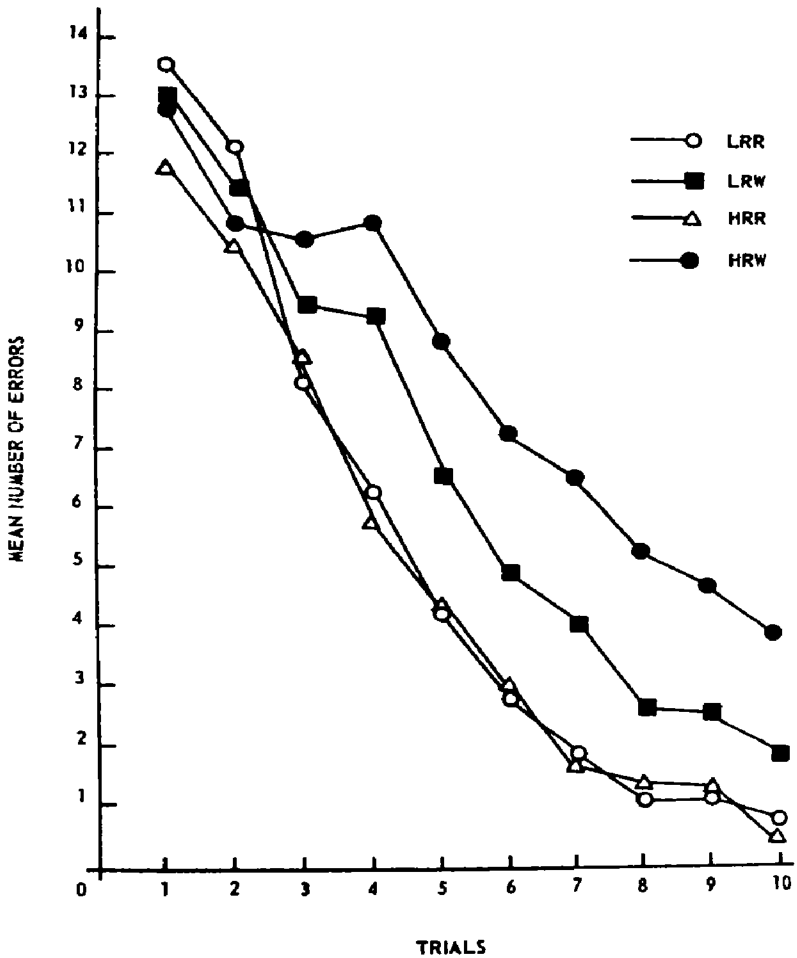


Figure 1. Mean errors per trial for the 4 differential pronunciation conditions (HRR and LRR = Rule 1; HRW and LRW = Rule 2)

was applied to the number of correct items recalled on the MFR task. Only the sequence of pronunciation effect was significant, $F(1, 72) = 9.64$, $p < .01$. Subjects for whom the sequence of pronunciation was W-R recalled an average of 10.41 items; those for whom the sequence was R-W recalled an average of 8.64. The absence of a significant direction of recall effect is in agreement with the results found by Sardello and Kausler (1967). The absence of a significant differential pronunciation effect is more difficult to interpret since in the Sardello and Kausler study overt pronunciation was compared to no pronunciation whereas overt pronunciation was required of all subjects in the present study. It may be that the degree of overt pronunciation (practice) was sufficiently high in each of the 4 groups to eliminate the positive relationship normally found between trials to criterion and MFR performance. The authors do not have a tenable explanation for the significant sequence of pronunciation effect The possibility that it is simply an artifact is suggested by the absence of such an effect in the previous 4 analyses.

DISCUSSION

The results of this experiment demonstrated that the use of Rule 2 has a severely retarding effect on the rate of VD acquisition. The difficulty of Rule 2 responding is perhaps best realized by examination of Figure 1. where performance curves of the 4 groups are shown. The nature of the interaction revealed in Figure 1 (i.e., facilitated acquisition followed by slow improvement across trials) is consistent with results of studies previously mentioned (Lovelace, 1966; Newby and Young, 1972; Smith and Jensen, 1971; Underwood and Freund, 1968, Experiment I; Underwood, Jesse, and Ekstrand, 1964). Invariably, these authors attributed the net negative performance of the "Rule 2" groups to the faster rate with which frequency units were added to the R item relative to the W item. The results of the present study cannot be similarly accounted for since frequency units accrued to the R and W items at an approximate constant rate throughout the task.

One possible explanation for the adverse effects of Rule 2 responding is that the subject must process more information than those subjects for whom the use of Rule 1 is functional. Newman et al. (1972) suggested that the application of Rule 2 (e.g., "A is more frequent than B, reject A, select B") might involve a greater number of operations, thus requiring more time than the application of Rule 1 (e.g., "A is more frequent than B, select A"). Support for such an explanation was reported by Underwood et al. (1964), who found that a brief anticipation interval

impeded performance of subjects using Rule 2 but had little effect on the use of Rule 1. This notion seems especially feasible in view of the identical intervals (1.5 seconds) used in this study and the Underwood et al. study.

A second explanation, or one which can be used in conjunction with the first, involves the method of feedback. More specifically, the presence of the R item could conceivably have been a source of interference in the HRW and LRW conditions. This hypothesis presented itself to us based on the observation that a number of subjects in the Rule 2 groups seemed to focus their attention on the R item during the feedback interval. It may be that under these conditions responding to the W item is quantitatively different (i.e., worth less) than responding to the R item. Or in terms of frequency theory, the presence of both items may have allowed for the simultaneous accrual of frequency to the R item (via RRs) and to the W item (via PRs). In either case the frequency difference would be decreased and the difficulty of the task increased. Although such a suggestion is presently without empirical support, it would appear to have testable characteristics.

Another unexpected finding was the generally inferior performance of the HRW group relative to the LRR and LRW groups. Also contrary to frequency theory, which predicts facilitated performance in the high-ratio groups relative to the low-ratio groups, were the data indicating a comparable rate of acquisition in the HRR and LRR groups. The finding of no significant differences between the HRR and LRR groups would appear to indicate that a portion of the study interval was devoted to implicit rehearsal of the R item in the latter group. Consequently, the assumption of implicit rehearsal must also be considered a possibility in the LRW group. If this assumption is valid, then Rule 1 would have been the appropriate mode of response in this group. The better performance of the LRR group compared to the LRW group would be expected on the basis of a greater frequency difference between the R and W items.

Notes

The authors would like to thank Barbara L. Gardecki for her critical reading of the initial draft of this paper. This experiment constitutes a portion of a thesis submitted by the first author in partial fulfillment of the requirements for the degree of Master of Arts at Central Michigan University, 1974. Peter M. Lisiecki is now at the Department of Psychology, Illinois Institute of Technology, Chicago, IL 60616. Portions of this paper were presented at the August 1975 meeting of the American Psychological Association, Chicago, Illinois. Requests for reprints should be sent to Terry M. Libkuman, Department of Psychology, Central Michigan University, Mt. Pleasant, MI 48859. Received for publication January 27, 1976; revision, February 14, 1977.

References

Ekstrand, B. R., Wallace, W. P., and Underwood, B. J. 1966. A frequency theo of verbal discrimination learning. *Psychological Review* 73:566–578.

Lovelace, E. A. 1966. Knowledge of correct and incorrect items in verbal d crimination. *Psychonomic Science* 6:363–364.

Newby, R. W., and Young, R. K. 1972. Presentation of correct and incorre items in verbal discrimination learning. *Journal of Experimental Psycholo* 96:471–472.

Newman, S. E., Suggs, R. E., and Averitt, C. H. 1974. Use of rule 1 and rule in verbal discrimination training. *Journal of Experimental Psychology* 10: 531–533.

Radtke, R. C., McHewitt, E., and Jacoby, L. 1970. Number of alternatives a rate of presentation in verbal discrimination learning. *Journal of Expe mental Psychology* 83:179–181.

Sardello, R. J., and Kausler, D. H. 1967. Associative recall in verbal discrimin tion learning as related to pronunciation and degree of practice. *Psych nomic Science* 8:253–254.

Smith, S., and Jensen, L. 1971. Test of the frequency theory of verbal discrim nation learning. *Journal of Experimental Psychology* 87:46–51.

Thorndike, E. L., and Lorge, I. 1944. *The teacher's word book of 30,000 wor* New York: Teachers College, Columbia University.

Underwood, B. J., and Freund, J. S. 1968. Two tests of a theory of verbal di crimination learning. *Canadian Journal of Psychology* 22:96–104.

Underwood, B. J., Jesse, F., and Ekstrand, B. R. 1964. Knowledge of rights ar wrongs in verbal discrimination learning. *Journal of Verbal Learning an Verbal Behavior* 3:183–186.

AMERICAN JOURNAL OF PSYCHOLOGY
December 1977, Vol. 90, No. 4, pp. 699–704

Number preferences in ratio estimation and constant-sum scaling

C. Plug
University of South Africa

Subjects employed in ratio estimation and constant-sum scaling show strong tendencies to use rounded numbers. The effect is more pronounced in constant-sum scaling and causes results obtained with this method to be slightly more variable.

In direct-scaling methods, subjects attempt to match numbers to the subjective effect of stimulation so that the numbers are proportional to the strengths of the corresponding sensations. Judging from the introspections reported by subjects and the variability of their number estimates this matching is not an easy task. Usually there is considerable uncertainty whether a particular stimulus produces an effect that is twice, 3 times, or 4 times that of another. Under these conditions subjects must choose a single value from a fairly wide range of numbers that all seem more or less equally applicable. The final choice may thus be expected to depend partially on factors other than the strength of sensation, an example being rounding tendencies. These consist of reporting mainly integer numbers and multiples of 5 and 10 (Stevens, 1956; Baird, Lewis and Romer, 1970). Although Stevens (1975, p. 286) described number preferences as "a source of error peculiar to magnitude estimation," it has been shown that similar preferences occur in ratio estimation (Baird, Lewis, and Romer, 1970). The purpose of this article is to compare the extent of these rounding tendencies in ratio estimation and constant sum scaling and to investigate their effect on results obtained by these two methods.

PROCEDURE

The experiment that provided the data to be discussed here formed part of a series of investigations relating to the psychophysics of form. Details are given elsewhere (Plug, in press). In this particular experiment, 20 "diamond" figures

(parallelograms presented standing on one corner) were used. Ten of the figures were elongated in a vertical direction, with height-to-breadth ratios of 1.30, 1.43, 1.58, 1.63, 1.85, 2.25, 2.56, 2.92, 3.66 and 4.75, while the other 10 were elongated horizontally, with height-to-breadth ratios equal to the inverses of the first 10. These stimuli were presented to 2 roughly equal groups of subjects (total $N = 50$) in an irregular order and its reverse. The subjects were required to estimate the height-to-breadth ratios of the stimuli, expressing their estimates directly as ratios. Two other groups of subjects (total $N = 50$) were asked to judge the same stimuli, but to express their estimates as a division of 100 points between the height and breadth of each figure, a procedure usually referred to as constant-sum scaling. Subjects were first-year university students.

RESULTS

In order to provide an overall view of number preferences, the 1,000 constant-sum estimates (the first of the 2 numbers) are presented in the form of a histogram in Figure 1. The stimulus ratios and the arithmetic mean of the estimates on each stimulus are also shown. It is clear that, except for the numbers 33 and 67, which represent ratios of .5 and 2, practically all estimates are multiples of 10 or 5. The figure leaves no doubt as to the extent of rounding in constant-sum scaling. In order to compare the number preferences in ratio estimation with those of the constant-sum method, the distribution of all estimated ratios is shown in Figure 2, which also indicates the geometric mean of the estimates for each stimulus. Again rounding tendencies are clearly present, although not as extensive as for constant-sum scaling. This could conceivably be due to the fact that ratios may be expressed in 2 different ways, namely

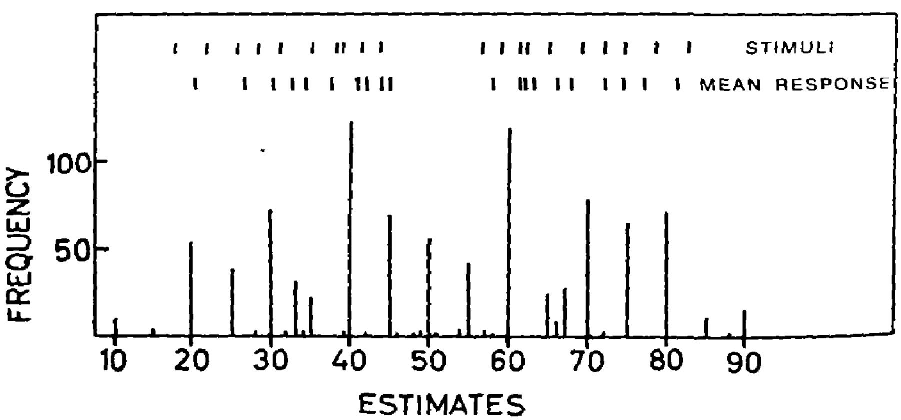


Figure 1. Distribution of estimates of 20 stimulus ratios by 50 subjects, expressed as constant-sum numbers; position of stimulus ratios and arithmetic means of the estimates on the constant-sum scale are indicated at the top of the figure

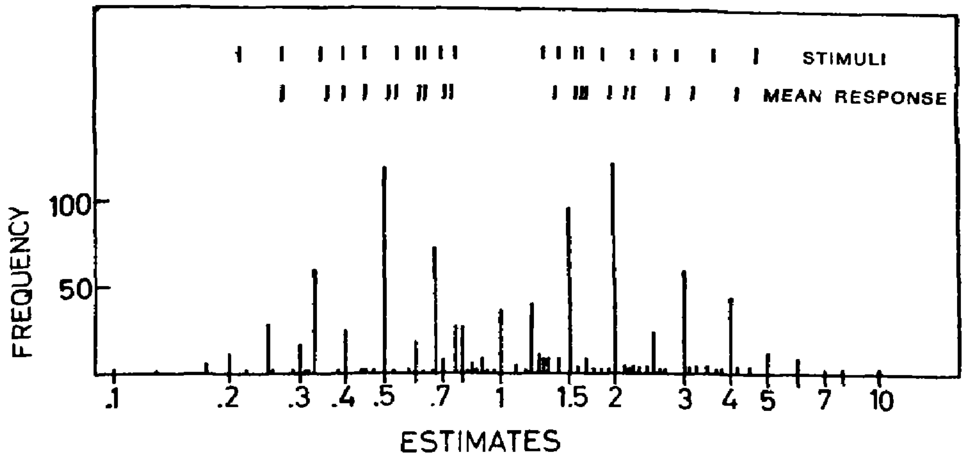


Figure 2. Distribution of direct estimates of 20 stimulus ratios by 50 subjects; stimulus ratios and geometric means of the estimates are indicated at the top of the figure

as decimals (e.g., 0.4) or fractions (e.g., ⅜). Subjects were allowed to use either or both of these response strategies. The results show that subjects who used only decimals ($N = 10$) each employed on the average 13.6 different values in the ratio-estimation task. The corresponding number for those who used fractions only ($N = 19$) was 11.3 and for those who used both decimals and fractions ($N = 21$) it was 11.0. Thus the difference in rounding tendencies between the 2 methods is clearly not due to the utilization of 2 intermixed modes of responding in ratio estimation.

The extent of rounding may be quantitatively measured by the amount of uncertainty associated with the distributions of the 50 responses for each stimulus (John, 1969). The response uncertainty U_i associated with stimulus i may be calculated by means of the formula:

$$U_i = -\sum_j p_j \cdot \log_2 p_j$$

where p_j is the proportion of times that response j is given to stimulus i. The maximum possible value of U_i will be $\log_2$ of the number of responses per stimulus. Since 50 responses were elicited for each stimulus, the maximum value of U_i in this case is 5.64 bits.

In the case of constant-sum scaling the average response uncertainty associated with each of the 20 stimuli was 2.46 bits, while for ratio estimation it was 2.83 bits. A t test showed the difference to be significant at the 1% level. Thus responses are spread more uniformly and over more different values in ratio estimation than in constant-sum scaling, indicat-

ing more severe rounding tendencies in the constant-sum method.

Are rounding tendencies equally severe for ratios above and below 1? Since 10 different stimulus ratios and their inverses were used in the present experiment, this could be directly tested. The response uncertainty relating to each stimulus was calculated separately for the 10 stimuli with ratios smaller than 1 and for the 10 with ratios larger than 1. For the constant-sum method the 2 mean values were 2.48 and 2.44 bits respectively, while for ratio estimation they were 2.86 and 2.80 bits. Neither of the differences is statistically significant, so that rounding tendencies seem to be roughly uniform over the whole range of stimulus ratios.

Having found that number preferences are more pronounced in constant-sum scaling than in ratio estimation, one must then consider briefly the effect that this may have on the mean and standard deviation of the estimates on each stimulus. This problem was considered by Baird, Lewis, and Romer (1970), who calculated the expected effects on the basis of some necessary assumptions. Given that the rounding unit is constant along the response continuum and assuming that the response value before rounding comes from a normally distributed population, it was shown that the maximum systematic effect of rounding on the mean response is .4 of the rounding unit. However, this applies only when the standard deviation of the responses is very small. The effect disappears completely if the standard deviation of the estimates exceeds approximately half of the rounding unit. The standard deviations of the constant-sum estimates obtained in this study were calculated for each of the 20 stimuli separately and were found to vary from 4.1 to 7.7 units on the constant-sum scale, with an average of 6.4. Thus the standard deviations are mostly larger than the common rounding unit of 5. No systematic errors due to rounding are therefore expected to be present in the averages of the constant-sum estimates. Since rounding is less severe in ratio estimation, the same conclusion applies to this method.

Rounding will, however, also affect the standard deviation of the estimates. The calculations of Baird, Lewis, and Romer (1970) indicate that the standard deviation will be slightly increased by rounding when the standard deviation and the rounding unit are of comparable size. Furthermore, for rounding units of up to 2 standard deviations, the effect increases monotonically with the size of the rounding unit. Since estimates are more evenly spread over the response continuum in ratio estimation than in the constant-sum method, the effect of rounding should be smaller in ratio estimation. Thus the relative increase of the standard deviation of estimates on any particular stimulus due to rounding is expected to be larger for the constant sum method than for ratio estimation.

In order to investigate this prediction the standard deviations of esti-iates in the 2 methods must be compared along a common continuum. 'he standard deviation of ratio estimates is usually calculated in terms f the logarithms of the ratios, since estimates on a single stimulus are ound to be approximately log-normally distributed. Since constant-sum stimates are also usually transformed to logarithmic ratios before scale alues are calculated, this seems to be an appropriate common continuum. ll estimates were therefore transformed to logarithmic ratios and the ariances of estimates relating to the 20 stimuli were calculated separately nd pooled. An F test indicates that the pooled variance for the constant-um method is significantly larger ($p < 1\%$) than that for ratio estima-ion, thus supporting the prediction.

Since the 2 methods were applied under identical conditions and to he same stimuli, there seems to be only one possible explanation for the greater variability of constant-sum estimates, namely, that it results from he more constrained use of numbers. This conclusion is not limited to he choice of a scale of logarithmic ratios as a common continuum. Trans-ormation of the ratio estimates to constant-sum numbers and the use of the constant-sum scale as a common continuum should lead to the same result. Calculations indicate that this is so. The variance in constant-sum scaling again significantly exceeds that in ratio estimation ($p < 5\%$) if all estimates are expressed as constant sum numbers.

DISCUSSION

The results presented above have shown that the tendency to use rounded numbers is not limited to magnitude and ratio estimation, but is particularly severe in constant sum scaling. Although it is unlikely to have a significant systematic effect on the mean estimates, it causes results ob-tained with this method to be more variable than those obtained with ratio estimation. Other considerations being equal, it may therefore be advisable to use direct ratio estimation instead of the constant-sum method if the number of subjects is severely limited.

Notes

Requests for offprints should be sent to C. Plug, Department of Psychology, University of South Africa, P.O. Box 392, Pretoria 0001, Republic of South Africa.
Received for publication August 13, 1976; revision, October 19, 1976.

References

Baird, J. C., Lewis, C., and Romer, D. 1970. Relative frequencies of numerica responses in ratio estimation. *Perception and Psychophysics* 8:358–362.

John, I. D. 1969. Stimulus discriminability in the magnitude estimation an category rating of loudness. *Perception and Psychophysics* 6:78–80.

Plug, C. In press. The psychophysics of form: Scaling the oblongness of plan figures. *South African Journal of Psychology*.

Stevens, S. S. 1956. The direct estimation of sensory magnitude — loudnes *American Journal of Psychology* 69:1–25.

Stevens, S. S. 1975. *Psychophysics: Introduction to its perceptual, neural an social prospects*. New York: Wiley.

AMERICAN JOURNAL OF PSYCHOLOGY
December 1977, Vol. 90, No. 4, pp. 705–707

Hans-Lukas Teuber: 1916–1977

Hans-Lukas Teuber, professor of psychology and head of the Department of Psychology at Massachusetts Institute of Technology, died January 4, 1977, while on vacation. He was born in Berlin, August 7, 1916, completed preparatory studies at the College Français there in 1934, and subsequently went to the University of Basel, 1935–39. He came to the United States in 1941, where he was given an assistantship at Harvard University for his first year and was appointed as a research psychologist at the Cabot Foundation, 1942–44 and 1946–47, while continuing his graduate study. He served as a Cabot Fund Fellow during the years 1944–46. Teuber received his Ph.D. in 1947 in Harvard's Department of Social Relations, under the supervision of Gordon Allport. He had served in the U.S. Armed Forces during the latter part of World War II. After obtaining his degree, he was appointed research associate of experimental psychology at New York University's Bellevue Medical Center, and in the Graduate School of Arts and Sciences. He remained at NYU, advancing in grade to full professor from 1947 to 1961. Teuber then was invited to MIT as Professor of Psychology and head of the Department of Psychology, where he served from 1961 until his death at the age of 60. He was married in August of 1941 to Marianne Liepe, who, with his sons Andreas and Christopher, survives him. This, in broad outlines, gives the chronology of his life. The richness of his influence requires a somewhat different kind of account.

Hans-Lukas Teuber's death marks the end of an era in physiological psychology. If one were to specify the beginning of this era one might choose the symposium on Brain Function in Behavior at the Penn State APA meeting of 1950. The symposium was chaired by Donald Hebb and the participants were Harry Harlow, Karl Pribram, Luke Teuber, and Joseph Zubin. It was agreed beforehand that no punches would be pulled and none were. The result was a most delightful exchange of views that had the packed audience rocking with laughter and repeated applause. Suddenly the relationship between brain and behavior had come to life once again after decades of "mass action" neuropsychology and "black box" behaviorism. The success of the symposium can be attributed to the fresh approach, skilled technique, thorough scholarship, dedicated commitment, and open friendship that each of the speakers brought to the symposium — attributes that were to characterize this small coterie of young Turks and their students for the next quarter century. Luke Teuber was a central figure in maintaining these high standards through periods of dearth and opulence, when spirits flagged and when the centrifugal forces of growing responsibility intruded. His contemporaries as well as his heirs were his students.

Teuber's contributions over the more than twenty-five years during which his leadership was exercised centered on the threefold relationship between brain

anatomy, brain pathology, and measurable behavior. Correlations were honed to new precision and much of the hearsay and mythology that had plagued observational clinical neurology were expunged.

However, this same period in scientific history also saw the development of instruments to study the electrical activity of the brain — the electroencephalogram, the elicitation of event-related potentials both stimulus and response evoked, the sophisticated use of microelectrodes extracellularly and intracellularly, and the use of computers to control stimulus and response and to analyze the complexities of the electrical signals recorded. Though he did not participate in the mainstream of these investigations, Teuber did not ignore them nor the important discoveries in neurochemistry that began to be made toward the end of his life. His highly successful basic psychology course at MIT and his public lectures at all times reflected his awareness of what was transpiring — an awareness often critical of developments, and conservative and demanding of solid proof of the value of a finding that might turn out to be just an intriguing novelty of only temporary importance.

Teuber's breadth as well as his conservatism reflects his history. With a basic European education he matriculated at Harvard to obtain a doctorate in *social* psychology. His thesis assayed the long-term effects of a variety of psychiatric treatments of delinquent adolescents and concluded that the treatments available at the time had no effect whatsoever. His distrust of strictly behavioral solutions to problems matured at that time. During the war he became associated with Morris Bender in the care of brain-injured patients. Bender's sharp critical acumen appealed to the wary Teuber — and the patients showed disturbances in behavior and experience that could be reliably measured and repeatedly observed. Earlier interests were rekindled. Teuber's father had preceded Wolfgang Köhler at Tenneriffe, where the observations on insightful behavior by chimpanzees were made. Also, Teuber had married the lovely daughter of a professor of the history of art and one of the founders of the "Gestalt" movement that Kohler, Kafka, and Wertheim later brought into psychology. Marianne Teuber, continuing her father's interests, made an excellent companion and collaborator throughout the years. She read critically everything Luke wrote and brought to his attention relevant material outside his own immediate sphere of inquiry. Her own scholarship surfaced again recently in calling the attention of the scientific community to the work of Escher, pointing out the significance for perception of the perspective transformations that Escher used to trick his viewers into seeing the impossible.

Although best known for his scholarly and research contributions in the field of perception, Teuber was preeminent in all of human neuropsychology. With Zangwill, Milner, Luria, and Hécaen, Teuber set the stage for a new era of thoughtful, quantitative behavioral observation of patients who have suffered brain injury. His influence was felt well beyond the published record. His warm friendship, combined with his devotion to excellence, provided nourishment over the whole range of the investigative effort.

Perhaps his crowning achievement was the establishment of a sound biologically oriented department of psychology at the Massachusetts Institute of Technology. Prior failures to make psychology "go" at MIT attest to the difficulties that had to be surmounted. In addition to his published documents and the influence on his students and colleagues, the extraordinarily gifted faculty of his de-

partment and the excellent data that they furnish their scientific colleagues give a measure of Hans-Lukas Teuber's judgment and dedication.

Personally, Lukas was a dear friend. Before I moved to the West Coast, it was the custom of our families to celebrate Thanksgiving together each year. My students became his students, his scientific problems became mine. During the recent decades distance interfered with custom, but we often walked the shores of the Pacific together, puzzling out a difference in views or charting a common purpose to right some administrative wrong. We were able to continue the openness of our early years. Our last appearance together was at a symposium of past presidents of the Physiological Division of the American Psychological Association. Luke gave a cautious endorsement of computers and their relevance to the analysis of possible mechanisms of brain function as well as of their usefulness in handling data. I remarked that if Luke had come around to espousing computers that there was hope for change in all of us no matter how "senior" we had become. Luke retorted with the warning that the blind use of instruments and overinstrumentation in general was the trap he had been avoiding, and he used as example the fact that the orientation selectivity of cells in the visual system was discovered by means of hand-held projectors, that mechanized projection of stimuli in Jung's laboratory in Freiburg had precluded this discovery. This is the type of lesson that we all learned so often and so fruitfully from Hans-Lukas Teuber. We shall miss him very, very much.

Karl H. Pribram, *Stanford University*

AMERICAN JOURNAL OF PSYCHOLOGY
December 1977, Vol. 90, No. 4, pp. 709–721

Book reviews

Carl P. Duncan
Northwestern University

The Psychology of Memory

By Alan D. Baddeley. New York: Basic Books, 1976. Pp. 430. $13.95.

The area of human learning and memory is undergoing a tremendous expansion in terms of popularity and paradigmatic approaches. Accompanying this growth has been a predictable upsurge in the number of "texts" on memory. Most of these have been focused either on efficiently and systematically summarizing the vast body of accumulated knowledge on how our memory functions or else on showing us a glimpse of the new horizons in cognitive research. However, there is a new book in the area — *The Psychology of Memory* by Alan D. Baddeley — that attempts to bridge the gap between the traditional Ebbinghaus approach and the emergent cognitive orientation first initiated by Bartlett In fact, Baddeley clearly points out in the beginning of the book that the current state of the literature reflects a tension between these two approaches, and interpretations of the research should be framed within this perspective.

Baddeley summarizes the literature with a specific bias, but he makes this clear from the outset: "In writing such a book, one can choose either to present the information and views as objectively as possible or to present a personal view. I have attempted the latter. Consequently, although I have made every effort to ensure that my descriptions of experiments and results are fair and accurate, my selection of topics, degree of emphasis, and theoretical interpretations are, I am sure, somewhat idiosyncratic" (p. xv). As a result, Baddeley's exposition of memory phenomena is more a crusade to gently persuade us to reorient our thinking about the subject than an attempt to present a totally comprehensive research summary. This tactic is effectively developed because Baddeley maintains a pervasive optimism and excitement for the future of the field. One is left with the impression that he is sincerely concerned with the direction the research is moving in, rather than with trying to leave his philosophical imprint upon it. There is more of the loving parent than the egotistical autocrat behind these suggestions.

There are four main techniques used by Baddeley in implementing this subjective approach. The first involves the content areas he covers. Besides the usual topics, such as the stages of information processing and mechanisms of forgetting, he includes a number of "nonstandard" topic areas, including the Freudian notion of forgetting and the Gestalt interpretation of memory mechanisms. In

addition, he takes a number of topics only briefly touched on in most previous textbook summaries and expands on them considerably. For instance, he has sections on motor memory, tactile memory, and olfactory memory. Furthermore, he gives extensive coverage to amnesia, state-dependent processes, drug effects, and mnemonists. Finally, he includes a chapter on the distinction between the Bartlett and Ebbinghaus orientations and how they have shaped our thinking. In all these efforts, it is most evident that Baddeley is trying to strip away our blinders with respect to what is appropriate and inappropriate in the study of memory.

A second manifestation of his unique perspective is in "editorial" comments which he interjects from time to time. In these, he attempts to point out certain problem areas that he has personally wrestled with and issues that seem to have been neglected or underdeveloped. As an example of the first type of comment, Baddeley notes that "as one whose visual imagery is subjectively quite vivid, I still find it surprising (social desirability notwithstanding) that such large differences in rated vividness of imagery should be accompanied by no apparent behavioral consequences" (p. 224). He then presents a short discussion of this topic and its relationship to the literature. Concerning the second type of comment, Baddeley begins his chapter on supernormal memory with the statement that "... far from being a distraction from the pursuit of understanding normal memory, such cases of supernormal memory provide not only clues as to the functioning of normal memory but also a way of testing the generality of our theories of memory" (p. 347–348). This is preliminary to a strong endorsement of this line of research.

The third way Baddeley makes us aware of his own personal views is through a series of rather pointed criticisms of theoretical positions he considers either poorly developed or overly restrictive of research activities. His most vehement objection is directed at the interference theory of forgetting. In the introduction, he confesses that he is locked in a love-hate relationship with interference theory. On the one hand, it has provided an exemplary model of a systematic attack on the factors that determine forgetting. However, he views the theory as being overly dull and intellectually stifling in terms of the future growth of research. Despite these opinions, he does include a chapter on interference theory that thoroughly covers its evolution and present positions.

There are several other theoretical positions that Baddeley deems deserving of some constructive criticism. Concerning Sternberg's serial exhaustive search model, he notes that "it is not easy to see what function an exhaustive recognition scan could have outside the limits of Sternberg's task ..." (p. 150). Using this as a springboard, Baddeley suggests that we have to examine our paradigms more carefully and ask ourselves whether we are simply playing laboratory games with our computers and reaction-time measurements or whether we are really contributing to a better understanding of how people really think and learn. Baddeley devotes several pages to enumerating the shortcomings of Atkinson and Shiffrin's conceptualization of short-term memory. As a final illustration, he cautions the reader about overapplying the Craik and Lockhart concept of levels-of-processing, because of the difficulty of obtaining an independent measure of exactly what the depth of processing is in a given situation. Although Baddeley criticizes a wide variety of positions, his main message is clear: paradigmatic "trendiness" is not healthy for memory research at this point in its development.

The final manifestation of Baddeley's personalized approach is in his concluding chapter, entitled "Where do we go from here?" In it, he discusses the directions we should be keeping in mind for the future evolution of memory investigations. Specifically, he notes that researchers have exhibited the unfortunate propensity to become preoccupied with fad areas, with the resultant danger that "...we lose track of what we have already accomplished and simply go round in circles, discovering and rediscovering the same phenomena" (p. 373). The remedy is for researchers to pay more attention to what Egon Brunswik called "ecological validity": the ability of findings to apply to phenomena outside the laboratory. His overall prognosis, however, is bright. He suggests that investigators have simply overextended their aspirations so that they do not fit the realities of the present situation.

The only major reservation I have concerning the book is specifying the audience to which it is directed. The book is not appropriate as a first overview of the field because of the subtle issues and criticisms Baddeley raises. Furthermore, its organization is atypical and difficult to follow without some background in the area. As Baddeley mentions in the introduction, he rejected the traditional organization — by input, storage, and retrieval of information — in favor of a more "chronological framework." This results in a topical fragmentation that I believe the novice would find difficult to follow. As an example, the concept of a short-term memory is covered in the distinction between the unitary versus dual process interpretations of memory (chapter 6), the paradigms and models used in the study of short-term retention (chapter 7), the structure of the working memory (chapter 8), the visual memory processes (chapter 9), and the auditory memory processes (chapter 10). With respect to its use as an advanced graduate text, the book lacks sufficient detail to sustain any in-depth analyses. As a result, it seems to fit best into an intermediate level — as a second undergraduate coverage of the material or an elementary graduate survey.

In general, the approach Baddeley uses is lively and refreshing. His style is quite readable, even conversational at times. What is more to his credit is that he does this while keeping an appropriately analytical perspective on the critical issues in memory research. He deftly handles theoretical minutiae without losing sight of the general direction of the topic under examination (which is excellently illustrated in his coverage of interference theory).

As one who is involved in the field of memory research, I found the book to be one of the most readable ones which I have encountered. It is obvious that Baddeley had an enjoyable time writing the book, which in turn makes it a delight to read. I believe that the "biases" he incorporates into the text make it more intellectually stimulating than the somewhat dryer texts that are the norm. Baddeley's wish, put forth in his introduction, was certainly realized: "The end result, though inevitably uneven and idiosyncratic, hopefully may still convey some of the richness and excitement that I myself derived from producing this very personal review . . ." (p. xvii).

Alan S. Brown, *Southern Methodist University*

Beyond Aesthetics — Investigations into the Nature of Visual Art
Edited by Don R. Brothwell. London: Thames and Hudson Ltd., 1976 (U.S. Distributor: Transatlantic Arts, Levittown, N.Y.) Pp. 212. $25.00.

"A scientific attitude to art is by no means incompatible with the nature of art, and the subject must be viewed at both the individual level, and in relation to the population at large — whether this is a preliterate community or an advanced technological one. We must see art both as an expression of some inner mechanisms of one individual in relation to his biology and environment, and at the same time as an aesthetic matter which deserves to have wide repercussions in any society . . ." (p. 17). Thus does the editor, who is also the author of two of the twelve scholarly chapters that comprise this interesting text, toss down his gauntlet — the study of visual art is too important to be left solely to those acknowledged "insiders" identified in the preface by C. H. Waddington of the Institute of Animal Genetics, Edinburgh. The art historians, art critics, and artists themselves concentrate their attention on the art object. Their studies of visual art must be supplemented with broadly defined, yet pertinent, behavioral observations and data if the aesthetic phenomena are to be more fully understood. *In short, emphasis is placed on the individual with the aesthetic experience*, rather than on the art object; psychologists, anthropologists, educators, biologists, and sociologists, among others, have contributions to make in what one might call "trans-aesthetic studies" of visual art.

Brothwell, who is with the Institute of Archaeology, University of London, presents his argument for this approach, and justification for the text, in the first chapter, "Towards a working definition of visual art." The essay is scholarly with nearly sixty relevant citations, yet brief with under seven pages of text, and remarkably readable and clear. The "outsiders" (psychologists, anthropologists, etc.) will probably agree with the arguments and approve of the approach for the most part, but "insiders" (art historians, etc.) will doubtless have many questions and less than unanimous approval of the approach. In this sense of its having little likelihood of conversion of readers to a viewpoint (a viewpoint to which the reviewer subscribes), the chapter alone must be judged less than compelling in its arguments. However, taken with the remainder of the text, the first chapter's impact might be differently judged, for the remaining chapters do generally attest to the contributions and potential contributions of the trans-aesthetic studies of visual art.

For example, in the third chapter, "Visual art, evolution and environment," Brothwell argues effectively that for humans the Darwinian rule includes the "survival of the *culturally fittest,*" which must, in turn, embrace the aesthetic aspects of the culture. He then reviews the development of visual art as related to cultural anthropology, emphasizing the point that the visual arts are sociocultural phenomena based on biological evolution and environmental adaptation, and in later stages influenced by societal and cultural changes. He concludes that art must now be viewed as a part of the science of human behavior. The typical psychologist is likely not to disagree with the conclusion, but he will have found the material leading to that conclusion generally foreign to his areas of expertise.

Most psychologists will feel at ease with chapter 2, "Primate perception and aesthetics," by Andrew Whiten of the Department of Experimental Psychology, University of Oxford, though some art "insiders" will be outraged at the very

suggestion that the "creations" of monkeys and apes can be treated as art objects. But the chapter does not deal with art objects; rather, it deals with the responses of primates in the aesthetic-behavior domain. Thus, after a very brief review of primate vision (the Old World apes and monkeys are similar to humans, e.g., with hue, saturation, and brightness discriminations, whereas the New World monkeys and lower primates have less well developed color vision), the chapter deals separately with "creation" (ape picture-making) and "appreciation" (aesthetic responses or preferences). The data in the first category are primarily based on the drawings of two chimpanzees, *Alpha* and *Congo*, although observations of a few other animals are included. There *are* clear indications of some rudimentary communality with human drawing — e.g., thematic elaboration, figure and space filling, balancing, figure completing, and differential treatment of multiple figures depending on their patterning. Preferences can be measured with use of operant techniques, and certain primate preferences are reported; e.g., four monkeys all showed color preference for blue, green, yellow, orange, and finally, red (humans prefer blue, red, green, violet, orange, and yellow, in that order). The author attempts to interpret such preferences and differences in preferences on a biological or evolutionary basis, and distinguishes two types of preference — pleasure or interest — that might be differentially involved as motivators for compositional control, calligraphic differentiation, thematic variation, optimum heterogeneity, and other characteristics of visual art.

M. D. Vernon of the University of Reading is the author of chapter 4, "The development of visual perception in children." She has written with her usual clarity, interest, thoroughness, and high scholarship. The chapter should be viewed as a brief, but classic, synopsis of perceptual development during the first few years of a child's life. The role of the environment in providing the experiential base (the familiar environment "makes sense") on which expectations are built is stressed, and the point is made that the child's capacity to perceive and understand is markedly reduced in unusual or unfamiliar environments. Thus, the child's perception of pictures is quite closely related to the degree of similarity of the picture to the environmental experiences of the child. Most psychologists will have little or no trouble in relating the data of perceptual development to a comprehensive study of visual art.

Two chapters deal with art education. Chapter 5, "Visual education for young children," by J. M. Pickering of Crewe and Alsager College of Higher Education, and chapter 11, "The artist and education in art and design in the sixties and the seventies," by Jean Creedy of Hastings College of Further Education, will be viewed by most psychologists as essays on curricula and methods of instruction, perhaps a bit too far removed from the behavioral data of the other chapters and from the behavioral (performance) goals set for the student either as creator or appreciator of visual art. Such a view is somewhat unjustified, for both chapters represent a movement in the direction of performance-based curricular design. For example, in the earlier chapter, Pickering draws on the generalizations of learning and perceptual theories, as well as observations of perceptual development, and applies these to requirements for visual learning situations in terms of space or shape, light, movement, and methods at the preschool, first- and middle-school levels. One of the conclusions should be noted: "The need for a child to find pleasure, excitement and interest in natural and man-made things as well as an ability to make aesthetic responses should be

of vital concern to all involved in education . . ." (p. 97). This conclusion points up a characteristic of both chapters that psychologists would probably consider a failing; namely, that the assertion is rhetorical rather than empirical, as are the chapters themselves for the most part. Arguments are given, improvements are called for, but little or no indications are given of methods and criteria, or even intentions for that matter, to evaluate the results at some later time. In fact, the later essay by Creedy seems overly involved in bureaucratic, degree-granting considerations that may be peculiar to contemporary art education in Britain. They seem to be of limited relevance to American psychologists, although ominous warnings (which the reviewer cannot fully appreciate) are given; viz., "Artists and designers in their present educational situations, as in their daily lives, have much to lose in an uncertain future, particularly with the onset of real economic stringency . . ." (p. 177), and "This is a moment for the greatest vigilence, not only on the part of those immediately concerned, but for any who claim genuine interest and understanding of the visual arts in society and its future overall cultural development . . ." (p. 178).

Two chapters deal specifically with vision and visual art: chapter 7, "Trompe l'oeil to rompe l'oeil: vision and art," by R. A. Weale of the Institute of Ophthalmology, University of London, and chapter 8, "Defective vision and art," by R. W. Pickford, Department of Psychology, University of Glasgow. Weale's essay will probably have a greater impact on "insiders" for it represents a selection of demonstrations of visual effects in art, a rearranged art-history-in-brief based on four "visually grammatical stages": (1) things are what they are, (2) things are what we know them to be, (3) things are what they appear to be, and (4) things aren't! For the "outsider," its value lies in its "explanation" of groupings of visual art as related to the domains of the viewer's perception and his visual sense. Pickford's essay is a review of work, some experimental, with an emphasis on the art object as influenced by, or *possibly* influenced by, visual defects of the artist who created it. Thus, the elongated figures of El Greco *may* have resulted from a possible astigmatic defect, since the elongation can be "corrected" with a −1 Diopter astigmatic lens at 15°; of course, it is also possible, as the author notes, that the artist had no visual defect, but rather intended to create the astigmatic effect in his paintings. Attention is given to other topics such as visual and haptic perception and art, art of the blind, and defective color vision.

Social-psychological and cultural influences are covered in two chapters: chapter 6, "Visual art: some perspectives from cross-cultural psychology," by Marshall H. Segall of the Psychology Department, Syracuse University, New York, and chapter 10, "Psychology, culture and visual art," by Pickford. Segall argues effectively that a work of art constitutes an aspect of a culture as well as a product of the behavior of the artist, and since the artist's behavior is doubtless influenced by the culture in which it is embedded, the art object may be studied by psychologists as an indicator of individual (artist's) behavior and of cultural influences on behavior. What Segall then presents is, in a sense, the other side of Weale's coin — i.e., an essay relating visual perception to art and the history of art from the psychological "outsider's" milieu. Segall's and Weale's chapters each benefit from the presence of the other, and from the contextual clarification provided by the inclusion of Pickford's second chapter. For here Pickford presents psychology and the psychology of art as most psychologists understand their field, even noting the beginning of experimental aesthetics with

Fechner in 1876. Here the Gestalt laws and other aspects of the psychology of perception are presented, albeit all too briefly, and related to the psychology of art and aesthetics. Individual differences (apperceptive types and mood types), preferences, cross-cultural comparisons, personality, and cultural changes are all touched on, but in only fourteen pages, including illustrations. Most psychologists will judge this chapter too brief and would be willing to forego *some* of the other material were that to permit an expansion of the psychological "stuff" with which Pickford deals.

The two remaining chapters are 9, "The artist in the population statistics," by Peter Cox of the United Kingdom's Government Actuary's Department, and 12, "Art, morals and Western society" by H. R. Rookmaaker, Faculty of Letters, Free University of Amsterdam. As Cox observes, "The statistics of artists may tell us little, or add nothing to other knowledge, about the population at large, but at least they can indicate something about the profession itself . . ." (p. 141), and, indeed, his analysis of data principally from Britain and the United States seems to validate his observation. The results are interesting and potentially useful; e.g., it is at least gratifying to learn that visual artists ". . . mostly work steadily, earn good money and do not lack an industrial patron . . ." (p. 150) for such has not always been the case during the past. The typical psychologist will be able to appreciate without difficulty the benefits (and dangers) of interpretations of the census-based statistical data, and he will be comfortable with Cox's essay. Rookmaaker's chapter is a different story, and it will be the "insider," especially the art historian, who will be more likely to appreciate and be at ease with the material which is behavioral only on the rhetorical level.

What is the sum? How does the text stand, as a text? It is interesting. It is scholarly. It may be important, for it deals with an important topic and seeks to stimulate some newer, behaviorally based research on the aesthetic phenomena especially as evidenced in visual art. To what extent will this text stimulate a movement toward "trans-aesthetic studies" of visual art; to what extent will it impact the behavioral fields (psychology, anthropology, sociology, etc.)? Probably less than the editor, authors, and reviewer would hope. What is its principal merit, then? Its merits are two: (1) what it covers, it covers extremely well. Each chapter represents the highest order of scholarship and expertise in its area. (2) It presents, and articulates clearly, an exciting idea: the idea that more might be learned and understood about aesthetics in general and the visual arts in particular if attention were turned from the art object itself to the behavioral aspects of the information source (the artist who creates the work) and the information receiver (the viewer who perceives it). Even for psychologists this is a change of emphasis and not merely a reaffirmation of the older tenets of experimental aesthetics. Persons with interests in the psychology of art will want to read this book; those with serious interests will want to refer to it.

Earl A. Alluisi, *Old Dominion University*

The Neuropsychology of Memory

By A. R. Luria. Translated by B. Haigh with an introduction by K. R. Pribram. New York: Wiley, 1976. Pp. 372. $22.95.

In this volume Luria draws upon his vast clinical experience to provide new insights into the nature of normal and pathological memory processes. As always, he has not restricted himself to reporting clinical data, although there is a great

deal of such data to be found here. Rather, he makes great efforts to integrate these data within a general theoretical framework of memory processes.

As is invariably the case with someone who has examined the problem in detail, Luria concludes that memory processes are very complex and diversified. He concludes the volume by pointing out that "This analysis has shown that human memory has a very complex structure. It is a grave mistake to regard it as a simple recording, storage, and retrieval of information. Memorizing is a highly complex process of analysis of incoming information, followed by its selection and coding. Recall is an equally complex process of choosing the necessary systems of connections from all the possible alternatives, performed by goal-directed mnemonic activity" (p. 345).

Luria provides insight into the nature of several dimensions that can be used to analyze the functional system of memory and the disturbances in this functional system. One of the most interesting of these dimensions distinguishes between modality-specific and modality-nonspecific disturbances. Pribram points out in his introduction that such a distinction has also proved useful in his work

In his analysis of modality-specific disturbances in memory, Luria focuses most of his attention on brain damage that disrupts verbal processing in one way or another but leaves other modalities intact. Of course this emphasis on disturbances in verbal processing comes as no surprise given his background in the study of aphasia.

In his analysis of some of these modality-specific disturbances concerned with verbal processing it would seem that the use of some distinction such as Tulving (1972) has proposed between episodic and semantic memory would be useful While this distinction has come under some criticism since Tulving originally proposed it, Luria does seem to be dealing with one thing when he talks about various difficulties patients with say, frontal-lobe lesions, have when recalling the information in a passage or list of words and with quite another thing when he talks about the inability of patients with local lesions of the parieto-occipital zones of the left hemisphere to grasp and retain certain logico-grammatical relations. It is not at all clear that any neat distinction such as that between episodic and semantic memory will ever be fully satisfactory in neuropsychological studies of memory, but some theoretical clarification in this area would be useful.

In many ways, Luria's treatment of modality-nonspecific disturbances in memory is most interesting because it relies heavily on the theoretical foundations of Soviet psychology and therefore provides a new perspective for the Western reader. This theoretical background is especially in evidence when Luria deals with memory disturbances correlated with lesions to the frontal lobes. When dealing with such disturbances, he approaches the problem from a perspective that emphasizes the disruption of purposive mnemonic activity. This emphasis on goal-directedness, purposefulness, etc. is an integral part of the theory of activity (*deyatel'nost'*) that is the cornerstone of contemporary Soviet psychology. Almost all the notions used in this theory were proposed by L. S. Vygotsky (1956, 1960, and in press), but the theory has been expanded and reorganized in many respects by A. N. Leont'ev (1959, 1975). Luria (1976a) has said that the theory of activity implicit in his work is virtually the same as the one developed by Leont'ev. It seems that such an approach is particularly appropriate when dealing with frontal-lobe lesions since it provides criteria capable of bringing order to the mass of otherwise uninterpretable observations.

The careful reader of this volume will note several other points where Luria's work reflects his Vygotskian heritage. For example, the notion of internal speech (e.g., p. 199) is used occasionally in his analysis. It plays a much more important role in his work concerned with speech production and comprehension (e.g., 1976b). Luria has also used the notion of indirect memorizing (I think this would have been better translated as "mediated recall") in his work. This is based on a procedure Vygotsky (in press) used in some very interesting ways in his work with children. Luria's use of the notion of "sense" of a fragment or story (e.g., on pp. 269–270) also comes from Vygotsky's (1962) distinction in *Thought and Language* between sense (*smysl*) and meaning (*znachenie*). This distinction continues to play an important role in Soviet psychology. Meaning (*znachenie*) is closely connected with relatively context-free linguistic distinctions and is perhaps closest to what would be called semantic representation in the West. Sense (*smysl*), on the other hand, has been defined within the framework of the theory of activity and is concerned with the way in which a particular linguistic unit (e.g., word or sentence) is understood by virtue of its having occurred in a particular action and therefore in a particular context. It would have been useful to the reader for Luria to have gone over the distinction between sense and meaning once again in this work since he uses it freely.

In general, Haigh has done an excellent job of translating this volume. The only serious criticism I would raise has to do with the distinction between sense and meaning. The English word "semantic" sometimes appears in the text where the adjective "sense" (with an explanatory footnote) should have been used. There are also other less important points one could raise about the translation. For example, as mentioned above, it would seem more appropriate to use "mediated recall" rather than "indirect memorizing." Points falling under this latter category of criticism are much less important than the first criticism since the Western reader can understand what is intended by depending on the associated clinical procedure Luria uses. In general, however, the reader will find that the book reads very smoothly.

Luria has succeeded in presenting a great deal of clinical material organized around a few central themes (e.g., the predominance of retroactive inhibition over trace decay as an explanatory mechanism) in his theoretical framework. The book will undoubtedly be of interest to clinicians, but it also deserves the attention of experimental psychologists and psycholinguists interested in studying human memory. The fact that Luria is developing a theoretical approach unfamiliar in the West as well as introducing new data should make the book of particular interest to those interested in new approaches to processes in human memory and how they are disrupted.

James V. Wertsch, *Northwestern University*

References

Leont'ev, A. N. 1959. *Problems in the development of mind* [in Russian]. Moscow: Academy of Pedagogical Sciences Press RSFSR.

Leont'ev, A. N. 1975. *Activity, consciousness, personality* [in Russian]. Leningrad: Political Literature Press.

Luria, A R. 1976a. Personal communication.

Luria, A. R. 1976b. *Basic problems of neurolinguistics.* The Hague: Mouton.

Tulving, E. 1972. Episodic and semantic memory. In *Organization of memory*, ed. E. Tulving and W. Donaldson. New York: Academic Press.
Vygotsky, L. S. 1956. *Selected psychological studies* [in Russian]. Moscow: Academy of Pedagogical Sciences Press RSFSR.
Vygotsky, L. S. 1960. *The development of higher mental functions* [in Russian] Moscow: Academy of Pedagogical Sciences Press RSFSR.
Vygotsky, L. S. 1962. *Thought and language*. Cambridge, Mass.: MIT Press.
Vygotky, L. S. In press. *Mind and society*. Cambridge, Mass.: Harvard University Press.

Sensation and Perception: An Integrated Approach

By Harold R. Schiffman. New York: Wiley, 1976. Pp. 434. $14.95.

Textbooks on sensation and perception traditionally begin with chapters on each of the sensory modalities and follow these with chapters on topics in perception. This format reflects, in part, the influence of Wundt, whose nineteenth-century "mental chemistry" viewed sensations as the elements of consciousness and perceptions as complexes built up from sensations; however, it also reflects the way in which modern specialists have come to define their own territories. These territories are usually restricted to either sensation or perception. Moreover, vision and audition usually get the lion's share of coverage. Thus textbooks on sensation and perception tend to consist of collections of briefer versions of specialized treatments. Schiffman's *Sensation and Perception: An Integrated Approach* is, despite its title, such a collection. Schiffman begins with a chapter on psychophysics. This is followed by a chapter on vestibular sensation, three chapters on audition, chapters on kinesthesis and cutaneous sensation, temperature and pain, taste, smell, and three chapters on vision. The final chapters deal with the perception of form, movement, and space; the constancies and illusions; the development of perception; perceptual motor coordination; and the perception of time.

This textbook provides a conventional, well-illustrated, usually competent, if dated, treatment of topics in sensation and perception. Unfortunately, there are errors that serve to illustrate how very difficult it is for any one person to write with authority on such a variety of topics. Since the same holds true for reviewers, this review will focus on the sensory chapters.

The treatments of audition and vision are the best executed as well as the most extensive. Nevertheless, the audition chapter contains an error that has become almost a classic in psychoacoustics (for a discussion see the preface of S. S. Stevens's *Psychophysics*, Wiley, 1975). In the first audition chapter, Schiffman provides an explanation of the decibel scale that clearly explains that sound energy (I) is proportional to the square of sound pressure (P): $I = kP^2$. In a later audition chapter, he confuses sound energy and sound pressure in a discussion of loudness. He says that Stevens found two different exponents, .3 and .67, for the power function relating loudness to the magnitude of the sound stimulus; however, the exponent of .3 resulted from relating loudness to sound energy (I) whereas the exponent of .67 resulted from relating loudness to sound pressure (P). A simple substitution of P^2 for I converts the exponent from .3 to .6. (The

difference between .6 and .67 represents a real revision of the sone scale by Stevens.)

The lower senses tend to receive not only less space but also less thoughtful treatment. For example, the chapter entitled "Somesthesis I: Kinesthesis and Cutaneous Sense" includes a discussion of the use of the skin as an information channel for the blind. Schiffman praises the ingenuity of the apparatus developed to convert visual stimuli into skin stimuli; however, attempts at this kind of sensory conversion have a long history of failures. Communication through the skin is limited by the sensory capabilities of the skin, not the ingenuity of the engineer. Another section in this chapter describes the Pacinian corpuscle as a joint receptor and completely ignores its role as a touch receptor.

The chapter "Temperature and Pain" includes a good description of the Melzack-Wall gate theory of pain, but neither the controversy this theory has aroused nor the evidence against the theory were discussed.

The chapters on taste and smell contain some competent sections, but on the whole reflect the usual neglect of these topics in general texts. The major weakness of the taste chapter is its reliance on questionable turn-of-the-century psychophysical data and its omission of modern data. For example, Schiffman's statement that sensitivity to bitter is greatest on the back of the tongue, sweet, on the tip, etc., comes from Hänig's work in Wundt's laboratory in 1901 (this research, like much other material in the text, is cited only via secondary sources and the actual investigator is not mentioned). Virginia Collings in "Human taste response as a function of locus of stimulation on the tongue and soft palate" (*Perception and Psychophysics,* 1974) not only has provided better data on this issue, but she has also pointed out that Hänig's work has been repeatedly distorted in secondary sources. Hänig actually found only small differences in thresholds for various tongue areas, and Collings confirmed this result. Not even mentioned is one of the most important findings in modern taste psychophysics —that cross adaptation is nearly perfect within a quality but doesn't occur across qualities. Those interested in taste in other species will be particularly disappointed; Schiffman's treatment omits some very ingenious animal psychophysics that suggests that taste worlds may be quite similar across species.

The chapter on smell, like taste, omits much of the modern psychophysical research. It is discouraging to find the Zwaardemaker olfactometer, which dates from 1888, illustrated and discussed in a 1976 text. Control of the stimulus is crucial in olfaction and highly sophisticated olfactometers have been in use for years. The olfaction chapter ends with a discussion of pheromones, the fascinating substances that provide a chemical communication link between members of insect species. Schiffman provides a discussion of the possibility of human pheromones; however, the interested reader would do well to read the critique by Beauchamp et al. of this idea. They doubt that the concept of pheromones can be productively extended to mammals, let alone man. (See G. K. Beauchamp, R. L. Doty, D. G. Moulton, and R. A. Mugford, "The pheromone concept in mammalian chemical communications: A critique," in *Mammalian Olfaction, Reproductive Processes and Behavior,* ed. R. L. Doty, 1976).

The perception chapters are particularly well illustrated and contain some excellent descriptions of phenomena. It is vexing that so many of the descriptions are concluded with statements to the effect that there are a variety of

alternative explanations. Some of those explanations would have been a welcome addition to the text.

A variety of spelling errors will irritate the careful reader (e.g., *limpoid* for *lipoid, Gymnesma sylvestre* for *Gymnema sylvestre, macrosomatic* for *macrosmatic*). The glossary contains some of these as well as more serious errors of fact (e.g., "sapid" is defined as "capable of being dissolved" when it actually means capable of being tasted). The glossary also contains a number of definitions restricted in odd ways. "Frequency" is defined as "a characterization of sound waves by the number of cycles or pressure changes completed in a second" and "acuity" is defined as "the ability to detect, resolve, and perceive fine details of a visual display." The definitions of these terms should not be restricted to single sensory modalities.

Schiffman's book is an indifferent text teetering on the edge of being a very good one. There are just enough errors and misplaced emphases to give it a strange out-of-focus quality but there are also enough excellent sections to make one regret the bad ones even more. Some careful revision could make a future edition of this text a useful introductory source for graduate students.

Linda M. Bartoshuk, *John B. Peirce Foundation* and *Yale University*

Psychological Aspects and Physiological Correlates of Work and Fatigue

Compiled and edited by Ernst Simonson and Philip C. Weiser. Springfield, Ill.: C. C. Thomas, 1976. Pp. 445. $38.50.

The "deteriorative" changes resulting from continued activity stand in sharp contrast to the "adaptive" changes that facilitate performance. These "deteriorative" changes, in health and in disease, can be studied in terms of a broad range of variables including performance: motor (with wide variations in the demand on energy expenditure and skill), perceptual, and intellectual; experiential processes; and changes within the organism and its systems. This parallels A. G. Bills's classification of *fatigue* into objective, subjective, and physiological (in *The Psychology of Efficiency,* 1943). In terms of disciplines, the study of the phenomena covered by the umbrella term "fatigue" falls into the areas of psychology, physiology, and pathology.

In accord with the latter classificatory scheme, the present book was conceived by its originator, Ernst Simonson, as the second volume of a collaborative trilogy on work and fatigue. The first volume, entitled *Physiology of Work Capacity and Fatigue,* appeared in print in 1971. The third volume, to be entitled *Exercise and Fatigue in Disease,* barely reached the planning stage when Professor Simonson died.

The volume on the physiology of work capacity and fatigue is concerned primarily with metabolic changes and the function of the cardio-respiratory system, but deals also with such topics as microscopic morphology of the central nervous system, cerebral blood flow, and pupillography, which are directly relevant to the subject matter of the present, second volume.

This volume is the result of the labors of fourteen authors, including four German scientists, psychologists and physiologists, who incorporated into their accounts European literature largely inaccessible to American psychologists because of the language barrier.

The book opens with a synthesizing review of the physiological background and a chapter on biophysical models of fatigue conceived as a decrement in work capacity (though not necessarily a decrement in performance). A large section is devoted to motor aspects. The section on sensory functions is dominated by concern with vision, considered in reference to visual fatigue and to the study of visual functions as an avenue to the assessment of "general fatigue."

The section on central processing deals with sustained alertness and readiness for action, and the processes of information acquisition and information processing. Two chapters deal with subjective symptoms of fatigue and the role of motivation. Two other chapters review the literature on age and performance, in general and with special reference to the operation of motor vehicles. The volume closes with the comments of a psychologist, S. H. Bartley, on the question, "What do we call fatigue?" and a brief epilogue in which a physiologist, Hans Schaefer, gives thought to an interpretation of fatigue as an important, multifaceted phenomenon that affects human achievement and, at times, man's very existence. In contrast to the prevailing optimism concerning the interpretation of the psychological symptoms of fatigue in terms of the underlying physiological processes, Schaefer is more reserved. In his view, physiologists accept only the general principle that "there cannot be anything in the consciousness of man which is not somehow related to biochemical-physical processes in the brain" (p. 415). When we try to identify the specific physiological correlates of specific psychological states and processes, we run into severe difficulties. While we can identify the somatic bases of some "psychological" events (such as sleep and wakefulness), the rich world of human experience, including the experiences of boredom, tiredness, weariness, and exhaustion, cannot — not yet — be translated into neurophysiological terms: Not only is the matrix of brain structures and brain processes incredibly complex, but this complexity is further increased by the existence of feedback loops which, at present, we are unable to model adequately. Fatigue has been and remains a challenge to our deeper understanding of the functioning of the human organism.

The volume is well documented in terms of data, figures and references. It will serve as a distillate of the widely scattered literature, indispensable to all concerned professionally with the subject matter. Does it achieve a synthesis? I believe not, even though the introductory chapter by P. C. Weiser and the discussion of biophysical models by O. H. Schmidt point in that direction. In view of the "state of the art," characterized by a pluralism of approaches and plagued by a multiplicity of unresolved definitional issues, a true synthesis is beyond reach at present.

Josef Brožek, *Lehigh University*

Books received

(The books listed here have not as yet been noted in our pages. Listing here does not preclude their later review, however.)

Abell, Richard G. *Own Your Own Life*. New York: David McKay, 1976. Pp 234. $9.95.

Adams, Arthur A. *Test Yourself*. New York: Hart, 1976. Paperback. $1.95.

Adams, Junius. *Drink to Your Health*. New York: Harper & Row, 1976. Pp. 141 $7.95.

Adams, Margaret. *Single Blessedness*. New York: Basic Books, 1976. Pp. 264 $12.50.

Adler, Gerhard, ed. *C. G. Jung Letters. Volume 2: 1951–1961*. Princeton, N.J. Princeton University Press, 1976. Pp. 716. $20.00.

Alahuhta, Eila. *On the Defects of Perception, Reasoning, and Spatial Orientation Ability in Linguistically Handicapped Children*. Helsinki, Finland: Finnish Academy of Science and Letters, 1976. Pp. 152.

Albanese, Anthony Gerald. *The Plantation School*. New York: Vantage Press, 1976. Pp. 285. $6.95.

Albou, Paul. *Besoins et Motivations Economiques*. Paris: Presses Universitaires de France, 1976. Pp. 194. Paperback.

Allen, George J., and others. *Community Psychology and the Schools*. New York Halsted, 1976. Pp. 208. $12.95.

Allman, Lawrence R., and Jaffe, Dennis T. *Readings in Abnormal Psychology Contemporary Perspectives*. New York: Harper & Row, 1976. Pp. 396. Paperback. $5.95.

Anderson, Robert P., and Halcomb, Charles G., eds. *Learning Disability/Minimal Brain Dysfunction Syndrome*. Springfield, Ill.: C. C. Thomas, 1976 Pp. 283. $16.95.

Anderson, T. W., and Sclove, Stanley L. *Introductory Statistical Analysis*. Boston: Houghton Mifflin, 1976. Pp. 499. $13.95.

Arasteh, A. Reza, and Arasteh, Josephine D. *Creativity in Human Development* New York: Wiley, 1976. Pp. 154.

Ardrey, Robert. *The Hunting Hypothesis*. New York: Atheneum, 1976. Pp. 242 $10.00.

Arieti, Silvano. *Creativity: The Magic Synthesis*. New York: Basic Books, 1976 Pp. 448. $15.95.

Arnold, Wilhelm. *Person, Charakter, Personlichkeit*. Munich: Olzog Verlag, 1975. Pp. 464.

Aronow, Edward, and Reznikoff, B. *Rorschach Content Interpretation*. New York: Grune and Stratton, 1976. Pp. 362. $19.50.

Aronson, Elliot. *The Social Animal*. 2d ed. San Francisco: W. H. Freeman, 1976. Pp. 336. Paperback. $4.95.

Atkins, W. R., and Nurnberg, H. George. *How to Meditate without Attending a TM Class*. New York: Crown, 1976. Pp. 136. $6.95.

Ault, Ruth. *Children's Cognitive Development*. New York: Oxford University Press, 1977. Pp. 193. $9.00.

Axelrod, Robert, ed. *Structure of Decision*. Princeton, N.J.: Princeton University Press, 1976. Pp. 404. $25.00.

Bailey, Ronald H. *Violence and Aggression*. Boston: Little, Brown, 1976. Pp. 176. Paperback. $2.95.

Balsiger, Dave, and Sellier, Charles E., Jr. *In Search of Noah's Ark*. Los Angeles: Sun Classic, 1976. Pp. 218. Paperback. $1.95.

Barber, Theodore X. *Advances in Altered States of Consciousness and Human Potentialities*. Vol. I. New York: Psychological Dimensions, 1976. Pp. 698. $20.00.

Barber, Theodore X. *Pitfalls in Human Research*. New York: Pergamon Press, 1976. Pp. 117. $6.95.

Barham, R. M., and Boersman, F. J. *Orienting Responses in a Selection of Cognitive Tasks*. The Netherlands: Academic Book Services Holland, 1975. Pp. 154.

Barker, Philip. *Basic Child Psychiatry*. Baltimore: University Park Press, 1976. Pp. 274. $14.50.

Bar-Tal, Daniel. *Prosocial Behaviors: Theory and Research*. New York: Halsted, 1976. Pp. 197. $12.50.

Bassin, Alexander; Bratter, T. E.; and Rachin, R. L., eds. *The Reality Therapy Reader*. New York: Harper & Row, 1976. Pp. 691. $15.00.

Bates, Elizabeth. *Language and Context: The Acquisition of Pragmatics*. New York: Academic Press, 1976. Pp. 375. $19.95.

Bateson, P. P. G., and Hinde, R. A. *Growing Points in Ethology*. New York: Cambridge University Press, 1976. Pp. 548. $38.50.

Beck, Aaron T. *Cognitive Therapy and the Emotional Disorders*. New York: International Universities Press, 1976. Pp. 356.

Beck, Joan. *Effective Parenting*. New York: Simon and Schuster, 1976. Pp. 380. $8.95.

Becker, Ernest. *The Structure of Evil*. New York: The Free Press, 1976. Pp. 430. Paperback $2.95.

Bennett, Spencer, and Bowers, David. *An Introduction to Multivariate Techniques for Social and Behavioural Sciences*. New York: Wiley, 1976. Pp. 156. $14.95.

Bennett, Vivo, and Clagett, Cricket. *1001 Ways to be Your Own Boss*. Englewood Cliffs, N.J.: Prentice-Hall, 1976. Pp. 247. $10.00.

Bergman, Ingmar. *Face to Face*. New York: Pantheon, 1976. Pp. 118. Paperback $1.95.

Berlinski, David. *On Systems Analysis*. Cambridge, Mass.: MIT Press, 1976. Pp. 186. $15.00.

Berry, John W. *Human Ecology and Cognitive Style: Comparative Studies in Cultural Adaptation*. New York: Wiley, 1976. Pp. 242. $16.50.

Biderman, Albert D., and Drury, Thomas F., eds. *Measuring Work Quality for Social Reporting*. New York: Halsted, 1976. Pp. 297.

Biehler, Robert F. *Child Development: An Introduction*. Boston: Houghton Mifflin, 1976. Pp. 657. $13.95.

Bijou, Sidney. *Child Development: The Basic Stage of Early Childhood*. Englewood Cliffs, N.J.: Prentice-Hall, 1976. Pp. 210. Paperback. $5.95.

Billig, M. *Social Psychology and Intergroup Relations*. New York: Academic Press, 1976. Pp. 428. $29.95.

Binder, Virginia; Binder, Arnold; and Rimland, Bernard. *Modern Therapies*. Englewood Cliffs, N.J.: Prentice-Hall, 1976. Pp. 230. $8.95.

Black, James A., and Champion, Dean J. *Methods and Issues in Social Research*. New York: Wiley, 1976. Pp. 445.

Blake, Kathryn. *The Mentally Retarded*. Englewood Cliffs, N.J.: Prentice-Hall, 1976. Pp. 403. $11.95.

Bloom, Benjamin S. *Human Characteristics and School Learning*. New York: McGraw-Hill, 1976. Pp. 284.

Blumenthal, Monica D.; Chadiha, Letha B.; Cole, Gerald A.; and Jayaratne, Toby Epstein. *More About Justifying Violence: Methodological Studies of Attitudes and Behavior*. Ann Arbor, Mich.: Institute for Social Research, 1976. Pp. 401. Paperback. $13.00.

Bockar, Joyce A. *Primer for the Nonmedical Psychotherapist*. New York: Spectrum, Pp. 131.

Bolton, Brian, ed. *Psychology of Deafness for Rehabilitation Counselors*. Baltimore: University Park Press, 1976. Pp. 156. $14.50.

Bower, Sharon Anthony, and Bower, Gordon. *Asserting Your Self*. Reading, Mass.: Addison-Wesley, 1976. Pp. 244. Paperback. $4.95.

Brehm, Sharon S. *The Application of Social Psychology to Clinical Practice*. New York: Halsted, 1976. Pp. 272.

Brenner, Charles. *Psychoanalytic Technique and Pyschic Conflict*. New York: International Universities Press, 1976. Pp. 221.

Brewer, Marilynn B., and Campbell, Donald T. *Ethnocentrism and Intergroup Attitudes: East African Evidence*. New York: Halsted, 1976. Pp. 218. $15.95.

Brown, Hugh. *Brain and Behavior*. New York: Oxford University Press, 1976 Pp. 413. $12.00.

Brownfield, Charles A. *The Brain Benders*. New York: Exposition, 1972. Pp. 218. $6.50.

Bruce, Robert D. *Lacandon Dream Symbolism. Volume 1. Dream Symbolism and Interpretation*. Mexico City: Ediciones Euroamericanas, 1976. Pp. 131. Paperback.

Bruner, Jerome S.; Jolly, Alison; and Sylva, Kathy. *Play: Its Role in Development and Evolution*. New York: Basic Books, 1976. Pp. 716. $20.00.

Bueno de Mesquita, Bruce. *Strategy, Risk and Personality in Coalition Politics: The Case of India*. New York: Cambridge University Press, 1976. Pp. 198. $17.50.

Bullough, Vern L. *Sexual Variance in Society and History*. New York: Science History Publications, 1976. Pp. 185. Paperback. $4.95.

Burton, Arthur. *What Makes Behavior Change Possible?* New York: Brunner-Mazel, 1976. Pp. 334.

Caldwell, Bettye M., and Ricciuti, Henry N., eds. *Child Development and Social Policy: Review of Child Development Research*. Vol. 3. Chicago: University of Chicago Press, 1976. Pp. 581.

Cammer, Leonard. *Freedom from Compulsion*. New York: Simon and Schuster, 1976. Pp. 271.

Campbell, Robert, ed. *The Enigma of the Mind*. Boston: Little, Brown, 1976. Pp. 176. Paperback. $2.95.

Canter, David, ed. *Environmental Interaction*. New York: International Universities, 1975. Pp. 374.

Carstairs, G. M., and Kapur, R. L. *The Great Universe of Kota: Stress, Change and Mental Disorder in an Indian Village*. Berkeley: University of California Press, 1976. Pp. 176. $12.50.

Carter, Hugh, and Glick, Paul C. *Marriage and Divorce: A Social and Economic Study*. Cambridge, Mass.: Harvard University Press, 1976. Pp. 508. $17.50.

Casey, Edward S. *Imagining*. Bloomington: Indiana University Press, 1976. Pp. 240. $12.50.

Chadwick-Jones, J. K. *Social Exchange Theory*. New York: Academic Press, 1976. Pp. 431.

Chapman, Tony, and Foot, Hugh, eds. *Humor and Laughter: Theory, Research and Applications*. New York: Wiley, 1976. Pp. 348. $21.95.

Claghorn, James L. *Successful Psychotherapy*. New York: Brunner-Mazel, 1976. Pp. 208.

Clark, Alfred W., ed. *Experimenting with Organizational Life: The Action Research Approach*. New York: Plenum, 1976. Pp. 259. $25.00.

Cohen, Allan R.; Fink, Stephen L.; Gadon, Herman; and Willits, Robin D. *Effective Behavior in Organizations*. Homewood, Ill.: Richard D. Irwin, 1976. Pp. 615. $14.50.

Cohen, Edmund D., *C. G. Jung and the Scientific Attitude*. Totowa, N.J.: Littlefield, Adams, 1976. Pp. 167. $2.95.

Cole, James K., and Arnold, William J., eds. *1975 Nebraska Symposium on Motivation*. Lincoln: University of Nebraska Press, 1976. Pp. 579. Paperback. $11.95.

Coleman, James C. *Abnormal Psychology and Modern Life*. 5th ed. Glenview, Ill.: Scott, Foresman, 1976. Pp. 816. $15.95.

Coleman, Vernon. *Everything You Want to Know About Aging*. New York: Gordon, 1976. Pp. 226. $9.95.

College Programs for Paraprofessionals. New York: Human Sciences Press, 1976. Pp. 135.

Combs, Arthur W.; Richards, Anne Cohen; and Richards, Fred. *Perceptual Psychology: A Humanistic Approach to the Study of Persons*. New York: Harper & Row, 1976. Pp. 492. Paperback. $9.95.

Contributions to Anthropology: Selected Papers of A. Irving Hallowell. Chicago: University of Chicago Press, 1976. Pp. 534. $24.00.

Costello, Charles G. *Anxiety and Depression: The Adaptive Emotions*. Montreal: McGill–Queen's University Press, 1976. Pp. 149. $10.00.

Costler, Sherwin B., and Guerra, Julio J. *Assertion Training* Champaign, Ill.: Research Press, 1976. Pp. 252. Paperback. $5.95.

Craig, Grace J. *Human Development*. Englewood Cliffs, N.J.: Prentice-Hall, 1976. Pp. 498. $12.95.

Craighead, W. Edward; Kazdin, Alan E.; and Mahoney, Michael J., eds. *Behavior Modification: Principles, Issues, and Applications*. Boston: Houghton Mifflin, 1976. Pp. 556. $13.95.

Creer, Thomas L., and Christian, Walter P. *Chronically Ill and Handicapped Children*. Champaign, Ill.: Research Press, 1976. Pp. 200. Paperback. $7.95.

Culbertson, James T. *Sensations, Memories and the Flow of Time*. Santa Margarita, Calif.: Cromwell Press, 1976. Pp. 189. Paperback. $8.00.

Cull, John G., and Hardy, Richard E. *Problems of Runaway Youth*. Springfield, Ill.: C. C. Thomas, 1976. Pp. 166. Paperback. $11.95.

Daniels, Victor, and Horowitz, Laurence J. *Being and Caring*. Palo Alto, Calif.: Mayfield, 1976. Pp. 388. Paperback. $6.95.

Davids, Anthony, ed. *Child Personality and Psychopathology*. New York: Wiley, 1976. Pp. 299. $18.95.

Davidson, Park O., ed. *The Behavioral Management of Anxiety, Depression and Pain*. New York: Brunner-Mazel, 1976. Pp. 197. $13.50.

DeArmond, Stephen J.; Fusco, Madeline M.; and Dewey, Maynard M. *Structure of the Human Brain*. New York: Oxford University Press, 1976. Pp. 181. Paperback. $9.95.

DeCato, Clifford M., and Wicks, Robert J. *Case Studies of the Clinical Interpretation of the Bender Gestalt Test*. Springfield, Ill.: C. C. Thomas, 1976. Pp. 146. $9.75.

Deci, Edward L. *Intrinsic Motivation*. New York: Plenum, 1976. Pp. 324. $14.95.

DeLuca, Anthony J. *Freud and Future Religious Experience*. New York: Philosophical Library, 1976. Pp. 263. $12.50.

Denenberg, Victor H. *Statistics and Experimental Design for Behavioral and Biological Researchers: An Introduction*. New York: Wiley, 1976. Pp. 344 $17.50.

Dengrove, Edward. *Hypnosis and Behavior Therapy*. Springfield, Ill.: C. C Thomas, 1976. Pp. 405. $26.75.

Dentler, Robert A., and Shapiro, Bernard J., eds. *Readings in Educational Psychology: Contemporary Perspectives. 1976–77 ed.* New York: Harper & Row, 1976. Pp. 360. Paperback. $5.95.

Despert, J. Louise. *The Inner Voices of Children*. New York: Simon & Schuster, 1976. Pp. 162. Paperback. $3.95.

Dethier, V. G. *The Hungry Fly*. Cambridge, Mass.: Harvard University Press, 1976. Pp. 489. $30.00.

Devereaux, George. *A Study of Abortion in Primitive Societies*. Rev. ed. New York: International Universities Press, 1976. Pp. 414. $15.00.

DeVito, Joseph A., ed. *Communication: Concepts and Processes*. Rev. ed. Englewood Cliffs, N.J.: Prentice-Hall, 1976. Pp. 329. Paperback.

DiCaprio, Micholas S. *The Good Life*. Englewood Cliffs, N.J.: Prentice-Hall, 1976. Pp. 210. Paperback. $2.95.

Diggins, Dean, and Huber, Jack. *The Human Personality*. Boston: Little, Brown, 1976. Pp. 332. Paperback. $6.95.

Doi, Takeo (William J. Tyler, trans.), *The Psychological World of Natsume Soseki*. Cambridge, Mass.: Harvard University Press, 1976. $15.00. Pp. 161

Dorken, Herbert, and associates. *The Professional Psychologist Today*. San Francisco: Jossey-Bass, 1976. $15.00. Pp. 394.

Douglas, Tom. *Groupwork Practice*. New York: International Universities Press, 1976. Pp. 217.

Drake, James A. *Teaching Critical Thinking*. Danville, Ill.: Interstate Printers, 1976. Pp. 237. Paperback. $8.75.

Droz, R., and Rahmy, M. *Understanding Piaget*. New York: International Universities Press, 1976. Pp. 227.

Dunnette, Marvin P., ed. *Handbook of Industrial and Organizational Psychology*. Chicago: Rand McNally, 1976. Pp. 1740. $45.00.

Edwards, Allen L. *An Introduction to Linear Regression and Correlation*. San Francisco: W. H. Freeman, 1976. Pp. 213. Paperback. $5.95.

Eisenstein, Edward M., ed. *Aneural Organisms in Neurobiology.* New York: Plenum, 1976. Pp. 145. $19.50.

Eissler, Ruth S.; Freud, Anna; Kris, Marianne; and Solnit, Albert J., eds. *The Psychoanalytic Study of the Child.* Volume 31. New Haven, Conn.: Yale University Press, 1976. Pp. 573. $20.00.

Ekstein, Rudolf, ed. *In Search of Love and Competence.* New York: Brunner-Mazel, 1976. Pp. 369. $12.50.

Eleftheriou, Basil E., and Sprott, Richard L., eds. *Hormonal Correlates of Behavior.* Vol. I: *A Lifespan View.* New York: Plenum, 1976. Pp. 439. $29.50.

Eleftheriou, Basil E., and Sprott, Richard L., eds. *Hormonal Correlates of Behavior.* Vol. II: *An Organismic View.* New York: Plenum, 1976. Pp. 806. $29.50.

Elkind, David. *Child Development and Education.* New York: Oxford University Press, 1976. Pp. 274. $10.99.

Elkins, Dov Peretz. *Glad to Be Me.* Englewood Cliffs, N.J.: Prentice-Hall, 1976. Pp. 138. $8.95.

Emery, F. E., and Trist, E. L. *Towards a Social Ecology.* New York: Plenum, 1976. Pp. 239. Paperback. $5.95.

Endler, Norman S., and Magnusson, David. *Interactional Psychology and Personality.* New York: Halsted, 1976. Pp. 663. $22.50.

Epileptic Seizures, Behavior, Pain. Baltimore: University Park Press, 1976. Pp. 372. $29.50.

Ernst, Ken. *Pre-Scription: A TA Look at Child Development.* Milbrae, Calif.: Celestial Arts, 1976. Pp. 127. Paperback. $4.50.

Essman, W. B., and Valzelli, L., eds. *Current Developments in Psychopharmacology.* Vol. 2. New York: Halsted, 1976. Pp. 266. $20.00.

Essman, W. B., and Valzelli, L., eds. *Current Developments in Psychopharmacology.* Vol. 3. New York: Wiley, 1976. Pp. 303. $25.00.

Estes, W. K., ed. *Handbook of Learning and Cognitive Processes.* Vol. 2. *Conditioning and Behavior Theory.* New York: Halsted, 1975. Pp. 373. $18.50.

Eysenck, H. J., and Wilson, G. D. *A Textbook of Human Psychology.* Baltimore: University Park Press, 1976. Pp. 343. $16.50.

Eysenck, H. J., and Wilson, Glenn. *Know Your Own Personality.* New York: Penguin Books, 1976. Pp. 206. Paperback. $1.95.

Family Life and Child Development. New York: Child Study Press, 1976. Pp. 45. Paperback. $2.50.

Farber, Leslie H. *Lying, and Other Topics.* New York: Basic Books, 1976. Pp. 232. $10.00.

Farina, Amerigo. *Abnormal Psychology.* Englewood Cliffs, N.J.: Prentice-Hall, 1976. Pp. 207. Paperback. $5.95.

Fein, Sylvia. *Heidi's Horse.* Pleasant Hill, Calif.: Exelrod Press, 1976. Pp. 160. Paperback. $15.00.

Feinstein, Barbara, and Cavanaugh, Catherine. *The New Volunteerism: A Community Connection.* Cambridge, Mass.: Schenkman, 1976. Pp. 178.

Feldman, M. Philip, and Broadhurst, Anne, eds. *Theoretical and Experimental Bases of the Behaviour Therapies.* New York: Wiley, 1976. Pp. 459. $26.00.

Fields, Sidney Jacob. *The Person Circle.* Hicksville, N.Y.: Exposition Press, 1976. Pp. 183. $8.50.

Finkel, Norman J. *Mental Illness and Health.* New York: Macmillan, 1976. Pp. 129. Paperback. $3.95.
Fisher, Seymour, and Fisher, Rhoda L. *What We Really Know About Child Rearing.* New York: Basic Books, 1976. Pp. 262. $12.50.
Flavel, John H. *Cognitive Development.* Englewood Cliffs, N.J.: Prentice-Hall, 1976. Pp. 286. Paperback.
Formanek, Ruct, and Gurian, Anita. *Charting Intellectual Development: A Practical Guide to Piagetian Tasks.* Springfield, Ill.: C. C. Thomas, 1976. Pp. 119. $10.50.
Foulkes, S. H. *Group-Analytic Psychotherapy: Method and Principles.* New York: Gordon and Breach, 1976. Pp. 177. $13.50.
Frank, Johann Peter. *A System of Complete Medical Policy.* Baltimore: Johns Hopkins University Press, 1976. Pp. 469. $18.50.
Frankel, Fred H. *Hypnosis: Trance as a Coping Mechanism.* New York: Plenum, 1976. Pp. 185. $14.95.
Freeman, Thomas. *Childhood Psychopathology and Adult Psychoses.* New York: International Universities Press, 1976. Pp. 293. $16.50.
Fried, Robert. *Introduction to Statistics.* New York: Wiley, 1976. Pp. 304. $12.95.
Friedberg, John. *Shock Treatment Is Not Good for Your Brain.* San Francisco Glide Publications, 1976. Pp. 190. Paperback. $6.95.
Friedrich, Otto. *Going Crazy: An Inquiry into Madness in Our Time.* New York: Simon & Schuster, 1976. Pp. 384. $9.95.
Fry, Ruth T., and Hall, Joyce. *The Symbolic Profile.* Houston, Tex.: Gulf Publishing Co., 1976. Pp. 85. $10.95.
Furstenberg, Frank F., Jr. *Unplanned Parenthood.* New York: Free Press, 1976 Pp. 293. $13.95.
Gallant, Donald M., and Simpson, George M., eds. *Depression: Behavioral, Biochemical, Diagnostic, and Treatment Concepts.* New York: Halsted, 1976 Pp. 351. $25.00.
Gambrill, Eileen, and Richey, Cheryl. *It's Up to You.* Millbrae, Calif.: Les Femmes, 1976. Pp. 168. Paperback. $4.95.
Gardiner, Muriel. *The Deadly Innocents.* New York: Basic Books, 1976. Pp 190. $8.95.
Gary, A. L., and Glover, B. *Eye Color, Sex and Children's Behavior.* Chicago: Nelson-Hall, 1976. Pp. 170. $10.00.
Gay, L. R. *Educational Research: Competencies for Analysis and Application* Columbus, Ohio: Merrill, 1976. Pp. 354. $12.95.
Gay, L. R. *Educational Research: Models for Analysis and Application.* Columbus, Ohio: Merrill, 1976. Pp. 158. Paperback. $4.95.
Gaylin, Willard. *Caring.* New York: Knopf, 1976. Pp. 199.
Gedo, John E., and Goldberg, Arnold. *Models of the Mind.* Chicago: University of Chicago Press, 1976. Pp. 220. Paperback. $4.45.
Gedo, John E., and Pollock, George H., eds. *Freud: The Fusion of Science and Humanism. Psychological Issues.* Vol. IX, Nos. 2/3, Monograph 34/35. New York: International Universities Press, 1976. Paperback. Pp. 447.
Gerard, Harold B., and Miller, Norman. *School Desegregation.* New York: Plenum, 1976. Pp. 315. $17.95.

Getman, Julius G.; Goldberg, Stephen B.; and Herman, Jeanne B. *Union Representation Elections: Law and Reality*. New York: Russell Sage Foundation, 1976. Pp. 218. $7.50.

Gibson, Janice T. *Psychology for the Classroom*. Englewood Cliffs, N.J.: Prentice-Hall, 1976. Pp. 563. Paperback. $9.95.

Gill, Merton M., and Holzman, Philip S., eds. *Psychology Versus Metapsychology: Psychoanalytic Essays in Memory of George S. Klein*. New York: International Universities Press, 1976. Pp. 376. $17.50.

Gilman, Sander L., ed. *The Face of Madness*. New York: Brunner-Mazel, 1976. Pp. 111. Paperback.

Glasser, William. *Positive Addiction*. New York: Harper & Row, 1976. Pp. 159. $7.95.

Glick, Robert A., et al. *Psychiatric Emergencies*. New York: Grune and Stratton, 1976. Pp. 300.

Goldman, Jacquelin. *Becoming a Psychotherapist*. Springfield, Ill.: C. C. Thomas, 1976. Pp. 129. $9.50.

Goodman, Libby, and Mann, Lester. *Learning Disabilities in the Secondary School*. New York: Grune and Stratton, 1976. Pp. 256. $16.50.

Gordon, Chad, and Johnson, Gayle, eds. *Readings in Human Sexuality: Contemporary Perspectives*. New York: Harper & Row. 1976. Pp. 317. $5.95.

Gorecki, Jan, ed. *Sociology and Jurisprudence of Leon Petrazycki*. Urbana: University of Illinois Press, 1976. Pp. 144. $6.95.

Gorsuch, Richard L., and Malony, H. Newton. *The Nature of Man*. Springfield, Ill.: C. C. Thomas, 1976. Pp. 224. $12.75.

Gottman, John, et al. *A Couple's Guide to Communication*. Champaign, Ill.: Research Press, 1976. Pp. 220. Paperback.

Greenberg, Harvey R. *The Movies on Your Mind*. New York: Saturday Review Press, 1976. Pp. 273. Paperback. $4.95.

Greenspoon, Joel. *The Sources of Behavior: Normal and Abnormal*. Monterey, Calif.: Brooks-Cole, 1976. Pp. 224. $8.95.

Gregory, Susan. *The Deaf Child and His Family*. New York: Halsted, 1976. Pp. 256. $15.00.

Groddeck, Georg. *The Book of the It*. New York: International Universities Press, 1976. Pp. 346.

Guerin, Philip J., Jr. *Family Therapy*. New York: Wiley, 1976. Pp. 553. $24.50.

Guilleminault, C., et al., eds. *Narcolepsy*. New York: Halsted, 1976. Pp. 689. $40.00.

Gundelach, Peter, and Redder, K. W. *Arbejdskraftens mobilitet 3: Omfanget*. (The Scope of Labor Mobility.) Copenhagen: Socialforskningsinstituttet (The Danish National Institute of Social Research), 1976. Pp. 110.

Guttman, Samuel A., ed. *Psychoanalysis: Selected Papers of Robert Waelder*. New York: International Universities Press, 1976. Pp. 709.

Guze, Samuel B. *Criminality and Psychiatric Disorders*. New York: Oxford University Press, 1976. Pp. 181. $9.50.

Hacker, Frederick J. *Crusaders, Criminals, Crazies*. New York: Norton, 1977. Pp. 376. $9.95.

Hall, John F. *Classical Conditioning and Instrumental Learning: A Contemporary Approach*. Philadelphia: Lippincott, 1976. Pp. 518.

Hallahan, Daniel P., and Kauffman, James M. *Introduction to Learning Disabilities*. Englewood Cliffs, N.J.: Prentice-Hall, 1976. Pp. 310. $10.95.

Halpern, Werner I., and Kissel, Stanley. *Human Resources for Troubled Children*. New York: Wiley, 1976. Pp. 263. $17.95.

Hamilton-Merritt, Jane. *A Meditator's Diary*. New York: Harper & Row, 1976. Pp. 155. $6.95.

Hampden-Turner, Charles, and May, Rollo, eds. *Is There an Answer to Death?* Englewood Cliffs, N.J.: Prentice-Hall, 1976. Pp. 212. $3.75.

Hare, A. Paul. *Handbook of Small Group Research*. New York: Free Press, 1976. Pp. 781. $19.95.

Hart, Joseph; Corriere, Richard; and Binder, Jerry. *Going Sane: An Introduction to Feeling Therapy*. New York: Jason Aronson, 1976. Pp. 472. $9.95.

Harvey, John H.; Ickes, William J.; and Kidd, Robert F., eds. *New Directions in Attribution Research*. Vol. I. New York: Wiley, 1976. Pp. 467. $19.95.

Heckel, Robert V., and Salzberg, H. C. *Group Psychotherapy*. Columbia: University of South Carolina Press, 1976. Pp. 223.

Heiman, Julia; LoPiccolo, Leslie; and LoPiccolo, Joseph. *Becoming Orgasmic A Sexual Growth Program for Women*. Englewood Cliffs, N.J.: Prentice-Hall, 1976. Pp. 219. $8.95.

Heise, David R. *Causal Analysis*. New York: Wiley, 1976. Pp. 301. $15.95.

Heiskanen, Keikki. *A Systems Theoretical Approach to the Analysis of Social Actions and Decision*. Helsinki: Finnish Academy of Sciences and Letters, 1976. Pp. 115. Paperback.

Helmering, Doris Wild. *Group Therapy: Who Needs It?* Millbrae, Calif.: Celestial Arts, 1976. Pp. 211. Paperback. $4.95.

Hergenhahn, B. R. *An Introduction to Theories of Learning*. Englewood Cliffs, N.J.: Prentice-Hall, 1976. Pp. 402. $12.95.

Hilberman, Elaine. *The Rape Victim*. New York: Basic Books, 1976. Pp. 105. $7.95.

Hilgard, Ernest R., and Hilgard, Josephine R. *Hypnosis in the Relief of Pain* Los Altos, Calif.: William Kaufmann, 1975. Pp. 262.

Hirsch, H., ed. *The Family*. Basel: S. Karger, 1975. Pp. 566. $67.75.

Hirschowitz, Ralph G., and Levy, Bernard, eds. *The Changing Mental Health Scene*. New York: Halsted, 1976. Pp. 382. $25.00.

Hofmann, Adele D.; Becker, R. D.; and Gabriel, H. Paul. *The Hospitalized Adolescent*. New York: Free Press, 1976. Pp. 249. $14.95.

Hogan, Robert. *Personality Theory: The Personological Tradition*. Englewood Cliffs, N.J.: Prentice-Hall, 1976. Pp. 223. Paperback. $5.95.

Hollander, Edwin P. *Principles and Methods of Social Psychology*. 3rd ed. New York: Oxford University Press, 1976. Pp. 612. $13.00.

Hollander, Edwin P., and Hunt, Raymond G., eds. *Current Perspectives in Social Psychology: Readings with Commentary*. 4th ed. New York: Oxford University Press, 1976. Pp. 590. Paperback. $6.00.

Hollander, Edwin P., and Gordon, Myra E. *Instructor's Manual*. [To accompany *Principles and Methods of Social Psychology* and *Current Perspectives in Social Psychology*.] New York: Oxford University Press, 1976. Pp. 160

Honig, Werner K., and Staddon, J. E. R. *Handbook of Operant Behavior*. Englewood Cliffs, N.J.: Prentice-Hall, 1977. Pp. 689. $39.95.

Hornstein, Harvey A. *Cruelty and Kindness: A New Look at Aggression and Altruism*. Englewood Cliffs, N.J.: Prentice-Hall, 1976. Pp. 154. $7.95.

Horowitz, Irving Louis. *Professing Sociology*. Abridged ed. Carbondale: Southern Illinois University Press, 1976. Pp. 232. Paperback. $2.95.

Horrocks, John E. *The Psychology of Adolescence*. Boston: Houghton Mifflin, 1976. Pp. 574. $13.95.

Howard, Marion. *Only Human: Teenage Pregnancy and Parenthood*. New York: Seabury Press, 1976. Pp. 261. $8.95.

Howe, Michael. *Learning in Infants and Young Children*. Stanford, Calif.: Stanford University Press, 1976. Pp. 216. $8.95.

Human and Anti-human Values in Children's Books. New York: Council on Interracial Books for Children, 1976. Pp. 280. $14.95.

Illich, Ivan. *Medical Nemesis: The Expropriation of Health*. New York: Pantheon, 1976. Pp. 294.

Il'yuchenok, R. Yu. *Pharmacology of Behavior and Memory*. New York: Wiley, 1976. Pp. 216. $22.50.

Impact ERA. Millbrae, Calif.: Les Femmes, 1976. Pp. 287. Paperback. $4.95.

Inbar, G. F., ed. *Signal Analysis and Pattern Recognition in Biomedical Engineering*. New York: Halsted, 1976. Pp. 324. $25.00.

Interviewer's Manual. Rev. ed. Ann Arbor, Mich.: Survey Research Center, 1976. Pp. 142. Paperback. $8.00.

Isaacson, Robert L., and Pribram, Karl H., eds. *The Hippocampus*. Vol. 1. *Structure and Development*. New York: Plenum, 1976. Pp. 418. $24.50.

Isaacson, Robert L., and Pribram, Karl H., eds. *The Hippocampus*. Vol. 2. *Neurophysiology and Behavior*. New York: Plenum, 1976. Pp. 445. $24.50.

Ismael, Cristina. *The Healing Environment*. Millbrae, Calif.: Celestial Arts, 1976. Pp. 206. Paperback. $4.95.

Itil, T. M.; Laudahn, G.; and Herrmann, W. M., eds. *Psychotropic Action of Hormones*. New York: Halsted, 1976. Pp. 248. $25.00.

Izard, Camille. *Neuropsychologie et Tabac*. Paris: Masson, 1976. Pp. 287. Paperback.

Izenberg, Gerald N. *The Existentialist Critique of Freud*. Princeton, N.J.: Princeton University Press, 1976. Pp. 354. $16.50.

Jacoby, Neil H. *The Business-Government Relationship: A Reassessment*. Pacific Palisades, Calif.: Goodyear Publishing, 1976. Pp. 184. Paperback.

Jakab, Irene, ed. *Transcultural Aspects of Psychiatric Art*. Basel: S. Karger, 1975. Pp. 293. $51.75.

Jaynes, Julian. *The Origin of Consciousness in the Breakdown of the Bicameral Mind*. Boston: Houghton Mifflin, 1977. Pp. 467. $12.95.

Janov, Arthur, and Holden, E. Michael. *Primal Man: The New Consciousness*. New York: Crowell, 1976. Pp. 532. $10.00.

Jervis, Robert. *Perception and Misperception in International Politics*. Princeton, N.J.: Princeton University Press, 1976. Pp. 445. $22.50.

Johnston, Robert B., and Magrab, Phyllis R., eds. *Developmental Disorders*. Baltimore: University Park Press, 1976. Pp. 532. $14.50.

Journal of Community Health. Vol. 1, No. 2, 1975. New York: Human Sciences Press.

Jung, Carl G. (R. F. C. Hull, trans.) *Critique of Psychoanalysis*. Princeton, N.J.: Princeton University Press, 1975. Pp. 259. $3.45.

Jung, C. G. *Psychological Types.* Princeton, N.J.: Princeton University Press 1976. Pp. 608. Paperback. $5.95.

Jung, C. G. *The Symbolic Life.* Princeton, N.J.: Princeton University Press 1977. Pp. 904. $22.50.

Jung, C. G. *Symbols of Transformation.* Princeton, N.J.: Princeton University Press, 1976. Pp. 557. Paperback. $5.95.

Jung, John, and Bailey, Joan H. *Contemporary Psychology Experiments: Adaptations for Laboratory.* 2d ed. New York: Wiley, 1976. Pp. 212. Paperback

Kagan, Fred; Harwood, Theresa; Rickels, Karl; Rudzik, Allan D.; and Sorer, Heinz, eds. *Hypnotics: Methods of Development and Evaluation.* New York: Spectrum, 1976. Pp. 432.

Kahn, Samuel. *Essays in Freudian Psychoanalysis.* New York: Philosophical Library, 1976. Pp. 159. $10.00.

Kahn, Samuel. *Essays in Prejudice.* New York: Philosophical Library, 1976. Pp. 119. $6.00.

Kalt, Neil C., and Zalkind, Sheldon S. *Urban Problems.* New York: Oxford University Press, 1976. Pp. 543. $13.00.

Kaplan, Helen Singer. *The Illustrated Manual of Sex Therapy.* New York A & W Visual Library, 1976. Pp. 181. Paperback. $6.95.

Karrer, Rathe, ed. *Developmental Psychophysiology of Mental Retardation* Springfield, Ill.: C. C. Thomas, 1976. Pp. 506. $26.75.

Kart, Cary S., and Manard, Barbara B., eds. *Aging in America: Readings in Social Gerontology.* Port Washington, N.Y.: Alfred, 1976. Pp. 598. Paperback. $8.95.

Kaufman, Barry Neil. *Son Rise.* New York: Harper & Row, 1976. Pp. 153. $6.95

Keleman, Stanley. *Living Your Dying.* New York: Random House, 1976 Pp. 156. Paperback. $3.95.

Kendler, Howard H. *Basic Psychology.* 3d ed. Menlo Park, Calif.: W. A. Benjamin, 1976. Pp. 798.

Kendon, Adam; Harris, Richard M.; and Key, Mary Ritchie, eds. *Organization of Behavior in Face-to-Face Interaction.* Chicago: Aldine, 1975. Pp. 509

Kiell, Norman. *Varieties of Sexual Experience: Psychosexuality in Literature* New York: International Universities Press, 1976. Pp. 753. $25.00.

Kinnie, Ernest J., and Kelley, Jacqueline. *The Magic Theatre: The Psychology of Adjustment.* New York: Harper & Row, 1976. Pp. 328. Paperback. $5.95.

Kipnis, David. *The Powerholders.* Chicago: University of Chicago Press, 1976 Pp. 230. $12.50.

Kirschner, Joseph. *Eight Ways to Control Others.* Chicago: Henry Regnery, 1976. Pp. 184. $7.95.

Klahr, David, and Wallace, J. G. *Cognitive Development.* New York: Halsted, 1976. Pp. 244. $14.95.

Klein, George S. *Psychoanalytic Theory: An Exploration of Essentials.* New York: International Universities Press, 1976. Pp. 330. $17.50.

Koch, Joanne, and Koch, Lew. *The Marriage Savers.* New York: Coward, McGann and Geoghegan, 1976. Pp. 288. $8.95.

Koestler, Arthur. *The Thirteenth Tribe.* New York: Random House, 1976. Pp 255. $8.95.

Kogan, Nathan. *Cognitive Styles in Infancy and Early Childhood.* New York: Wiley, 1976. Pp. 146. $12.95.

Korchin, Sheldon J. *Modern Clinical Psychology*. New York: Basic Books, 1976. Pp. 672. $16.95.

Kornetsky, Conan. *Pharmacology: Drugs Affecting Behavior*. New York: Wiley, 1976. Pp. 275. $18.95.

Krebs, Dennis, ed. *Readings in Social Psychology: Contemporary Perspectives*. 1976–77 ed. New York: Harper & Row, 1976. Pp. 343. Paperback. $5.95.

Kreilkamp, Thomas. *The Corrosion of the Self*. New York: New York University Press, 1976. Pp. 235. $12.50.

Kreitler, Hans, and Kreitler, Shulamith. *Cognitive Orientation and Behavior*. New York: Springer, 1977. Pp. 447. $27.50.

Krippner, Stanley, and Villoldo, Alberto. *The Realms of Healing*. Millbrae, Calif.: Celestial Arts, 1976. Pp. 336. Paperback. $6.95.

Kuo, Zing-Yang. *The Dynamics of Behavior Development: An Epigenetic View*. New enlarged edition prepared by Gilbert Gottlieb. New York: Plenum, 1976. Pp. 237. $11.95.

Kupfer, David J.; Levine, Michael S.; and Nelson, John. *Mental Health Informations Systems*. New York: Marcel Dekker, 1976. Pp. 168. $19.50.

Kurtz, Ron, and Prestera, Hector. *The Body Reveals*. New York: Harper & Row, 1976. Pp. 148. Paperback. $4.95.

Laing, R. D. *The Facts of Life*. New York: Pantheon Books, 1976. Pp. 153. $7.95.

Lamb, Michael E. *The Role of the Father in Child Development*. New York: Wiley, 1976. Pp. 407. $17.95.

Lana, Robert E. *The Foundations of Psychological Theory*. New York: Halsted, 1976. Pp. 177. $12.95.

Lange, Arthur J., and Jakubowski, Patricia. *Responsible Assertive Behavior*. Champaign, Ill.: Research Press, 1976. Pp. 348. Paperback. $6.95.

Lantz, Herman R. *Marital Incompatibility and Social Change in Early America*. Beverly Hills, Calif.: Sage Publications, 1976. Pp. 48. Paperback. $3.00.

Laplanche, Jean. (Jeffrey Mehlman, trans.) *Life and Death in Psychoanalysis*. Baltimore: Johns Hopkins University Press, 1976. Pp. 148.

Laramy, John E. *Organizational Behavior in Schools Differing in Architectural Openness*. Danville, Ill.: Interstate, 1976. Pp. 60. Paperback. $1.95.

Lefcourt, Herbert M. *Locus of Control*. New York: Halsted, 1976. Pp. 211. $12.95.

Lehane, Stephen. *Help Your Baby Learn*. Englewood Cliffs, N.J.: Prentice-Hall, 1976. Pp. 295. $8.95.

Leichter, Hope Jensen. *The Family as Educator*. New York: Teachers College Press, 1976. Pp. 129. $7.00.

Leisman, Gerald. *Basic Visual Processes and Learning Disability*. Springfield, Ill.: C. C. Thomas, 1976. Pp. 439. $27.50.

Leitenberg, Harold, ed., *Handbook of Behavior Modification and Behavior Therapy*. Englewood Cliffs, N.J.: Prentice-Hall, 1976. Pp. 671. $32.00.

Leogrande, Ernest. *Second Chance to Live: The Suicide Syndrome*. New York: DaCapo Press, 1976. Paperback. $6.95.

Leslie, Gerald R. *The Family in Social Context*. 3d. ed. New York: Oxford University Press, 1976. Pp. 815. $13.00.

Levi, Lennart, ed. *Society, Stress and Disease*. Vol. 2. *Childhood and Adolescence*. New York: Oxford University Press. Pp. 551. $67.50.

Levin, Lowell S.; Katz, Alfred H.; and Holst, Erik. *Self Care*. New York: Prodist, 1976. Pp. 133. Paperback. $3.95.

Lewis, Jerry M.; Beavers, W. Robert; Gossett, John T.; and Phillips, Virginia Austin. *No Single Thread: Psychological Health in Family Systems*. New York: Brunner-Mazel, 1975. Pp. 260. $13.95.

Lezak, Muriel D. *Neuropsychological Assessment*. New York: Oxford University Press, 1976. Pp. 549. $16.95.

Lidz, Theodore. *The Person: His and Her Development throughout the Life Cycle*. New York: Basic Books, 1976. Pp. 615. $10.95.

Lief, Harold I., and Karlen, Arno. *Sex Education in Medicine*. New York: Halsted, 1976. Pp. 155. $12.95.

Lifton, Robert Jay. *Death in Life*. New York: Simon & Schuster, 1967. Pp. 594. Paperback. $5.95.

Lifton, Robert Jay. *Boundaries: Psychological Man in Revolution*. New York· Simon & Schuster, 1976. Pp. 113. Paperback. $2.95.

Lifton, Robert J. *The Life of the Self*. New York: Simon & Schuster, 1976. Pp. 190. Paperback. $2.95.

Lilly, John, and Lilly, Antonietta. *The Dyadic Cyclone*. New York: Simon & Schuster, 1976. Pp. 287. $8.95.

Lindgren, Henry Clay. *Educational Psychology in the Classroom*. 5th ed. New York: Wiley, 1976. Pp. 521.

Lindgren, Henry Clay, and Fisk, Leonard W., Jr. *Psychology of Personal Development*. 3d ed. New York: Wiley, 1976. Pp. 420. Paperback.

Lipsitt, Lewis P. *Developmental Psychobiology*. New York: Wiley, 1976. Pp. 143 $9.95.

Livingston, W. K. *Pain Mechanisms*. New York: Plenum, 1976. Pp. 250. $14.95

Lloyd, Barbara, and Archer, John, eds. *Exploring Sex Differences*. New York Academic Press, 1976. Pp. 280. $11.00.

Loevinger, Jane, *Ego Development*. San Francisco: Jossey-Bass, 1976. Pp. 504 $17.50.

Longchenpa (Herbert V. Guenther, trans.). *Kindly Bent to Ease Us*. Part II: *Meditation*. Emeryville, Calif.: Dharma, 1976. $10.95. Paperback. Pp 126 $4.95.

Lowen, Alexander. *Bioenergenetics*. Baltimore, Md.: Penguin Books, 1976. Pp 352. Paperback. $2.50.

Lugo, James O., and Hershey, Gerald L. *Handbook for Living Psychology*. 2d ed. New York: Macmillan, 1976. Pp. 242.

Lugo, James O., and Hershey, Gerald L. *Living Psychology*. 2d ed. New York: Macmillan, 1976. Pp. 640.

Lundsgaarde, Henry P. *Murder in Space City*. New York: Oxford University Press, 1977. Pp. 269.

Luthe, Wolfgang. *Creativity Mobilization Technique*. New York: Grune & Stratton, 1976. Pp. 303. $28.50.

Madden, Dehis, and Lion, Jonn R. *Rage, Hate, Assault and Other Forms of Violence*. New York: Halsted, 1976. Pp. 265. $20.00.

Magoon, Robert A., and Garrison, Karl C. *Educational Psychology: An Integrated View. 2nd ed.* Columbus, Ohio: Merrill, 1976. Pp. 483. $12.95.

Mahoney, Michael J. *Scientists as Subject: The Psychological Imperative*. Cambridge, Mass.: Ballinger, 1976. Pp. 249.

Mahoney, Michael J., and Mahoney, Kathryn. *Permanent Weight Control*. New York: Norton, 1976. Pp. 177. $7.95.

Malan, D. H. *A Study in Brief Psychotherapy*. New York: Plenum, 1976. Pp. 312. Paperback.

Malan, D. H. *The Frontier of Brief Psychotherapy*. New York: Plenum, 1976. Pp. 373. $19.95.

Malan, D. H. *Toward the Validation of Dynamic Psychotherapy*. New York: Plenum, 1976. Pp. 298. $19.50.

Malmquist, Eve. *Educational Research in Europe Today and Tomorrow*. Lund, Sweden: C. W. K. Gleerup, 1976. Pp. 442. Paperback.

Mandell, Arnold J. *The Nightmare Season*. New York: Random House, 1976. Pp. 216. $7.95.

Manning, Peter K., and Zucker, Martine. *The Sociology of Mental Health and Illness*. Indianapolis, Ind.: Bobbs-Merrill, 1976. Pp. 115. Paperback. $3.95.

Martin, Del. *Battered Wives*. San Francisco: Glide Publications, 1976. Pp. 288. Paperback. $6.95.

Martin, Peter A. *A Marital Therapy Manual*. New York: Brunner-Mazel, 1976. Pp. 206. $12.50.

Marty, Martin E. *A Nation of Behavers*. Chicago: University of Chicago Press, 1976. Pp. 239. $8.95.

Marx, Melvin H. *Introduction to Psychology*. New York: Macmillan, 1976. Pp. 753. $14.95.

Mash, Eric J.; Hamerlynck, Leo A.; and Handy, Lee C., eds. *Behavior Modification and Families*. New York: Brunner-Mazel, 1976. Pp. 362. $13.50.

Mash, Eric J.; Handy, Lee C.; and Hamerlynck, Leo A., eds. *Behavior Modification Approaches to Parenting*. New York: Brunner-Mazel, 1976. Pp. 254. $13.50.

Massaro, Dominic W. *Experimental Psychology and Information Processing*. Chicago: Rand McNally, 1975. Pp. 651.

Masserman, Jules, ed. *Social Psychiatry*. Vol. II. *The Range of Normal in Human Behavior*. New York: Grune & Stratton, 1976. Pp. 161. $16.50.

Masterton, R. B.; Hodos, William; and Jerison, Harry. *Evolution, Brain and Behavior*. New York: Wiley, 1976. Pp. 276. $14.95.

Matson, Floyd W. *The Idea of Man*. New York: Delta, 1976. Pp. 249. Paperback. $2.95.

Mayer, Richard E. *Thinking and Problem Solving: An Introduction to Human Cognition and Learning*. Glenview, Ill.: Scott, Foresman, 1976. Pp. 214. Paperback. $6.50.

Mazer, Nilton. *People and Predicaments*. Cambridge, Mass.: Harvard University Press, 1976. Pp. 279. $12.50.

McCubbin, Hamilton I.; Dahl, Barbara B.; and Hunter, Edna J., eds. *Families in the Military System*. Beverly Hills, Calif.: Sage Publications, 1976. Pp. 400. $20.00.

McDonnell, Kilian. *Charismatic Renewal and the Churches*. New York: Seabury, 1976. Pp. 202.

McGlashan, Alan. *Savage and Beautiful Country: The Secret Life of the Mind*. New York: Stonehill, 1976. Pp. 157. Paperback. $3.95.

McKeachie, W. J.; Doyle, C. L.; and Moffett, M. M. *Psychology*. 3rd ed. Reading, Mass.: Addison-Wesley, 1976.

McKinney, Fred; Lorion, Raymond P.; and Zax, Melvin. *Effective Behavior and Human Development.* New York: Macmillan, 1976. Pp. 504. $11.95.

McMahon, Frank B. *Abnormal Behavior: Psychology's View.* Englewood Cliffs, N.J.: Prentice-Hall, 1976. Pp. 478. $13.95.

Mead, D. Eugene. *Six Approaches to Child Rearing.* Provo, Utah: Brigham Young University Press, 1976. Pp. 141. Paperback. $4.95.

Mead, Margaret, and Heyman, Ken. *World Enough: Rethinking the Future.* Boston: Little, Brown, 1976. Pp. 218. $17.50.

Medin, Douglas L.; Roberts, William A.; and Davis, Roger T. *Processes of Animal Memory.* New York: Wiley, 1976. Pp. 267. $14.95.

Medinnus, Gene R., and Johnson, Ronald C. *Child and Adolescent Psychology.* 2d ed. New York: Wiley, 1976. Pp. 553. $13.95.

Mehrabian, Albert. *Public Places and Private Spaces.* New York: Basic Books, 1976. Pp. 354. $15.95.

Meites, Joseph; Donovan, Bernard T.; and McCann, Samuel M., eds. *Pioneers in Neuroendocrinology.* New York: Plenum, 1976. Pp. 327. $22.50.

Meltzer, H., and Wickert, Frederic R. *Humanizing Organizational Behavior.* Springfield, Ill.: C. C. Thomas, 1976. Pp. 438. $29.75.

Michel, Donald E. *Music Therapy.* Springfield, Ill.: C. C. Thomas, 1976. Pp. 134. $9.75.

Michelino. *The Keys to Truth.* New York: Philosophical Library, 1976. Pp. 90.

Miles, C. G., ed. *Experimentation in Controlled Environment.* Toronto: Addiction Research Foundation, 1976. Pp. 170. $9.00.

Miller, Kenneth M., ed. *Psychological Testing in Personnel Assessment.* New York: Halsted, 1976. Pp. 192. $15.95.

Miller, Kent S. *Managing Madness.* New York: Macmillan, 1976. Pp. 185. $9.95.

Miller, William R., and Munoz, Richardo. *How to Control Your Drinking* Englewood Cliffs, N.J.: Prentice-Hall, 1976. Pp. 246. Paperback. $3.95.

Monty, Richard A., and Senders, John W., eds. *Eye Movements and Psychological Processes.* New York: Wiley, 1976. Pp. 550. $35.00.

Moos, Rudolf H., ed. *Human Adaptation: Coping With Life Crises.* Lexington, Mass.: D. C. Heath, 1976. Pp. 446. Paperback. $5.95.

Moriarty, Alice E., and Toussieng, Povl W. *Adolescent Coping.* New York: Grune & Stratton, 1976. Pp. 220. $14.50.

Morris, Charles. *Psychology.* Englewood Cliffs, N.J.: Prentice-Hall, 1976. Pp. 650. $12.95.

Morris, Herbert. *On Guilt and Innocence.* Berkeley, Calif.: University of California Press, 1976. Pp. 171. $8.75.

Morris, Kenneth T., and Cinnamon, Kenneth M., eds. *Controversial Issues in Human Relations Training Groups.* Springfield, Ill.: C. C. Thomas, 1975. Pp. 137. $9.75.

Morris, Richard J. *Behavior Modification with Children.* Cambridge, Mass.: Winthrop Publishers, 1976. Pp. 229. $8.95.

Mostofsky, David I. *Behavior Control and Modification of Physiological Activity.* Englewood Cliffs, N.J.: Prentice-Hall, 1976. Pp. 504. $17.95.

continued on page 576

ontributors to Volume 90

ok review page numbers are italicized.

lams, Jack A., *336*
dridge, James W., 475
luisi, Earl A., *712*
ima, James K., *353*
neave, Fred, 549
rtoshuk, Linda M., *718*
ttig, William F., *165*
rman, Karen, 431
ke, Randolph, *539*
oko, Douglas J., 85, 577
linger, Deborah M., 145
dley, Drake R., 253
mer, Christine D., 645
ndley, Robin, 73
wn, Alan S., *168*
wn, L. B., 645
wn, Sam C., 157
žek, Joseph, *720*
nts, David C., 263
ssell, Edward M., 145
tino, Robert J., 577
appetta, Wayne, 577
ewski, William A., 219
ier, Alexis C., 195
n, Stanley, 609
ux, Kay, *350*
erman, Douglas K., 45
bs, Allen R., 309
can, Carl P., *168, 173,*
'8, *182, 326, 337, 347,*
'8, *349, 350*
s, James D., 511
ar, Paul, 551
ell, Michael T., 475
uson, Eva, *333*
erty, Charles F., 219
as, Joseph P., 635
rd, James L., 499
china, Joseph J., 291
l, Lois M., 517
l, William J., 517
Gerald, 431
s, Joan S., 609
on, William C., 231
berg, Michael, 599
Greenberg, Seth N., 675
Halff, Henry M., 383
Hasher, Lynn, 599
Herzberger, Sharon D., *174*
Herzog, Thomas R., 461
Hicks, Robert E., 95, 431
Hiew, Chok C., 207
Hilgard, Ernest R., *327*
Hinderliter, Charles F., 407
Hintzman, Douglas L., 489
Hogan, David E., 3
Holyoak, Keith J., *533*
Huesmann, L. Rowell, *357*
Jenkins, James J., 645
Kausler, Donald H., 481
Keating, Daniel P., *345*
Kestner, Jane, 29
Kimmel, H. D., 319, 529
Kleim, David M., 481
Lappin, Joseph F., *328*
Lembo, Vincent L., 17
Libkuman, Terry M., 689
Lippman, Louis G., 363
Lisiecki, Peter M., 689
Maki, Ruth H., 281
Manelis, Leon, 269
Mayzner, Mark S., *541*
McCullers, John C., 103
Mershon, Donald H., 17
Miller, George W., 431
Mistler-Lachman, Janet L., *323*
Mueller, John H., 157, 263
Murdock, Bennet B., Jr., *329*
Nation, Jack R., 103
Neely, James H., 447
Newell, Allan, *171*
O'Bannon, R. Michael, 55
Okuda, Arlene, 95
Pais, Nanette, 291
Parker, Donald E., *182*
Pavlik, W. B., 195
Pellegrino, James W., 419
Petrich, Judith, 419
Petry, Heywood M., 253
Pittenger, John B., 645
Platnick, Daniel M., 133
Plug, C., 699
Porac, Clare, 609
Pribram, Karl H., 705
Regan, Joan, 313
Riccio, David C., 407
Richards, Larry G., 63, 133
Routtenberg, Aryeh, *544*
Runcie, Dennis, 55
Runquist, Willard, *340*
Scheirer, C. James, 39, 565
Schiffman, H. R., 85, 577
Schmitt, John C., 39, 565
Sgro, Joseph A., 243
Shaughnessy, John J., 587
Sneed, Pamela R., 263
Sommer, Robert, 527
Stern, Leonard D., 489
Stober, Stephen R., 145
Swart, Dan, 299
Syme, G. J., 621
Syme, Lesley A., 621
Teghtsoonian, Robert, *338*
Thomas, John C., 499
Thompson, Richard W., 363
Thomsen, Dianne, 95
Tolkmitt, Frank J., 73
Tweney, Ryan D., 299
Walter, Donald A., 29, 655
Ward, Lawrence M., 609
Warren, Rik, 645
Waugh, Nancy C., 499
Weintraub, Daniel J., 461
Weir, C., 121
Weeks, James R., Jr., 291
Wertsch, James V., 715
Wong, Paul T. P., 663
Yadrick, Robert M., 481
Zechmeister, Eugene B., *536*
Zentall, Thomas R., 3

In memoriam

Hans-Lukas Teuber, 705

new from california

Emotion, Thought, and Therapy

A Study of Hume and Spinoza and the Relationship of Philosophical Theories to the Emotions of Psychological Theories of Therapy

Jerome Neu

Drawing upon modern discussion in philosophy of mind, psychology, and anthropology, Neu argues that the dependence of emotions on thought may help to explain the efficacy of psychoanalytic therapy and the possibilities that exist for psychological development.

256 pages, $12.95

Kinship and Marriage in Burma

A Cultural and Psychodynamic Analysis

Melford E. Spiro

Spiro's discussion of marriage and kinship in Burma is one of the few book-length treatments of these topics for any of the dominant ethnic groups in Southeast Asia.

336 pages, 8 drawings, 3 charts, $15.00

At bookstores

University of California Press Berkeley 94720